Protein Unfolding and Refolding by Single-Molecule Nanopore Analysis



A thesis submitted for the degree of

Doctor of Philosophy in Chemical Biology

Jianfei Feng

Department of Chemistry

UNIVERSITY OF OXFORD

March 2020

For my country, China

Acknowledgement

The work presented in this DPhil thesis would not have been possible without the close association with many people. I take this opportunity to extend my sincere gratitude and appreciation to all those who has contributed to this work.

First and foremost, I would like to thank my supervisor Professor Hagan Bayley for offering me the chance to study at the University of Oxford under his supervision. His constructive guidance, inspirational ideas and constant encouragement have lightened my way to accomplish this work. I greatly appreciate all his continued support that has made my DPhil experience intellectually stimulating. His energetic organization of various science meetings, including the group meetings every week, the Chemical Biology seminar every month and the one-on-one meeting every quarter, allow us to promote communication skills and broaden the horizon in frontline research.

His instant feedback has navigated my journey throughout the last five years. I am grateful and honored for the time I have spent to study under his supervision, which has been profoundly influential for my academic career.

I would also like to acknowledge my colleagues. Their ideas, discussion and assistance helped me to solve many research problems to get this work off the ground. When I first started in the laboratory, it was Dr. Michael Booth who spent a great deal of time bringing me up to speed in the lab and has been a great mentor over the course of my DPhil. In addition, he and Dr. Leon Harrington proofread this thesis and gave me valuable feedback. Dr. David Rodriguez-Larrea established an exemplary model system of co-translocational protein unfolding through a nanopore in the previous studies and provided technical advice in this work. He also provided me a thioredoxin variant. Ellina Mikhailova made the α-hemolysin pores that I have used throughout my time in the laboratory.

I am very grateful to Professor Mark Howarth for providing the streptavidin and its variants that were essential to my work, Dr. Jonathan Hopper and Professor Carol Robinson for the mass spectra of the oligonucleotide-protein conjugate, Dr. Jeffrey Tang of the Molecular Devices for the assistance with the electrophysiology data acquisition software. I also thank the Mechanical Workshop of the University of Oxford Chemistry Department for manufacturing the chambers for single-channel recordings.

I would like to thank Dr. Joongoo Lee, Dr. Gokce Su Pulcu, Dr. Linna Zhou and Pablo Martin-Baniandres for their helpful suggestion and discussion. I would like to thank my examiners, Professor Tom Brown and Professor Sergi

Garcia-Manyes, for taking the time to read and discuss this thesis.

I would like to express my sincere gratitude to the China Scholarship Council for my graduate scholarship.

Lastly, but just as importantly, my parents' love and unconditional support has always been my strength. I must be very grateful to their dedication and constant encouragement.

Jianfei Feng

University of Oxford

March 2020

Declaration

I declare that no part of this thesis has been accepted, or is currently being submitted, for any degree of diploma or certificate or any other qualification in this university or elsewhere. Unless acknowledged and stated otherwise in the text, the work contained in this thesis is my own, and is not the outcome of work done in collaboration.

Jianfei Feng

University of Oxford

March 2020

Abstract

Protein unfolding and refolding by single-molecule nanopore analysis

Nanoscale pores, nanopores, offer single molecule analytical capability that is achieved by electrophoretically driving the analytes in solution through them. The translocation of biological polymers through the nanopores has been investigated over three decades with particular focus on nucleic acids and, later on, peptides and proteins.

Chapter 1 presents the principles and technical aspects of nanopore stochastic sensing with an emphasis on a biological pore, α-hemolysin. It serves as a versatile model for sensing the translocation of nucleic acid polymers and discriminating stationary oligonucleotides at single-base resolution. With the development of the enzymatic ratcheting of polynucleotides through the pore, base-by-base, nanopore-based DNA

strand-sequencing has proven achievable. Protein information can also be read upon the protein analyte unfolding and translocating through the pore. This was achieved by either using a molecular motor or an electric potential acting on an oligonucleotide attached to the protein terminus.

Taking advantage of the ability of nanopores to discriminate stationary oligonucleotides, the potential-driven method for protein translocation can be advanced to determine the directionality of entry (N or C terminus-first). To achieve the discrimination, the protein is expected to attach a different oligonucleotide to each terminus. In Chapter 2, we used a site-specific N-terminal cysteine condensation reaction, followed by a disulfide bond forming reaction, to conjugate two different oligonucleotides to a model protein (thioredoxin), one to the N terminus and the other to the C terminus.

Chapter 3 presents the nanopore analysis of protein translocation direction. The oligonucleotide-attached protein bearing a single streptavidin bound at one of the two biotinylated oligonucleotide ends became unfolded as the unoccupied end was pulled through the pore under an applied potential. The different extent to which N- and C-terminal oligonucleotides block the pore indicated the direction of protein translocation through the pore.

Interrogation of the reversal process, protein refolding, is described in Chapter 4. Upon binding the two biotinylated oligonucleotides to streptavidin at both ends, a single molecule of thioredoxin was captured within the pore. The protein could be subjected to repeated cycles of translocation and retro-translocation. After each unfolding and retro-translocation back, it was allowed to refold for various delay times before translocation again. Therefore, this

system allows the characterization of protein folding at the single-molecule level. Thioredoxin collapses to a stable native-like structure rapidly, but the full refolding to its native state is approximately four orders of magnitude slower, possibly due to a rate-limiting step of *cis* prolyl peptide bond formation.

The full details of methods are listed in Chapter 5.

Contents

Acknowledgement ..... **i**

Declaration ..... **iv**

Abstract ..... **v**

Abbreviations ..... **xii**

Nomenclature ..... **xiv**

Chapter 1. General introduction **1**

1.1 Nanopore stochastic sensing 2

1.1.1 Principles of nanopore stochastic sensing 2

1.1.2 Nanopores 5

1.1.3 α-hemolysin pore 6

1.2 Analysis of nucleic acids 8

1.2.1 Detection of polynucleotides translocation 8

1.2.2 Identification of mononucleotides by the nanopore equipped with a molecular adapter 11
1.2.3 Single-base discrimination in stationary polynucleotides 12
1.2.4 Ratcheting of polynucleotide through the nanopore 16
1.3 Analysis of protein unfolding and translocation 20
1.3.1 Translocation of peptide chains 20
1.3.2 Translocation of unfolded proteins 21
1.3.3 Unfoldase-assisted co-translocational protein unfolding 24
1.3.4 Potential-driven co-translocational protein unfolding 28
1.4 Objective 32
Chapter 2. Bioconjugation of oligonucleotide to protein 33
2.1. Introduction 33
2.2. Results 37
2.2.1 Synthesis of crosslinker 37
2.2.2 Bioconjugation of oligonucleotide to protein 38
2.3 Discussion 45
Chapter 3. Discriminating the direction of co-translocational protein unfolding 46
3.1. Introduction 46
3.2 Results 52
3.2.1 Co-translocational protein unfolding pulled from N or C terminus 52
3.2.2 Discriminating the direction of protein translocation by oligonucleotide leader sequence 58
3.2.3 Discriminating the direction of protein translocation by oligonucleotide follower sequence 63

3.3 Discussion....72

Chapter 4. Single-molecule measurement of protein folding kinetics with a nanopore 78

4.1. Introduction 78

4.2 Results....83

4.2.1 Capture of a single molecule in the nanopore with traptavidin 83

4.2.2 Continuous unfolding and refolding of a single molecule 86

4.2.3 Describing folding kinetics of thioredoxin with a nanopore 90

4.3 Discussion....110

Chapter 5. Methods....114

5.1 General methods 114

5.2 Bioconjugation of oligonucleotide to protein....116

5.2.1 Instrumentation and sample analysis 116

5.2.2 Synthesis of crosslinker....118

5.2.3 Oligonucleotide modification....118

5.2.4 Protein expression and purification 119

5.2.5 Bioconjugation of oligonucleotide to protein 120

5.3 Single-channel recordings 121

5.3.1 αHL pore 121

5.3.2 Chamber....121

5.3.3 Single-channel recordings....124

5.3.4 Data analysis....129

Conclusion 130

Bibliography....132

Abbreviations

αHL	alpha-haemolysin
A/D convertor	analogue-to-digital convertor
ATP	adenosine triphosphate
btn	biotin
CBT	6-amino-2-cyanobenzothiazole
CsgG	Curlin sigma S-dependent growth subunit G
DNA	deoxyribonucleic acid
ddH_2O	distilled, deionized water
DMSO	dimethyl sulfoxide
DPhPC	1,2-diphytanoyl-*sn*-glycero-3-phosphocholine
EDTA	ethylenediaminetetraacetic acid
ESI	electrospray ionization
FPLC	fast protein liquid chromatography

HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid
HFIP	hexafluoroisopropanol
HPLC	high performance liquid chromatography
LB	lysogeny broth
mtHsp70	mitochondrial heat shock protein 70
MalE	maltose-binding protein
MS	mass spectrometry
MspA	*Mycobacterium smegmatis* porin A
NMR	nuclear magnetic resonance
oligo	oligonucleotide
PAGE	polyacrylamide gel electrophoresis
PMPI	*p*-maleimidophenyl isocyanate
poly	polynucleotide
RNA	ribonucleic acid
SDS	sodium dodecyl sulphate
TLC	thin-layer chromatography
TCEP	tris(2-carboxyethyl)phosphine
TEA	triethylamine
TIM	translocase of the outer membrane
TOM	translocase of the inner membrane
Tris	tris(hydroxymethyl)aminomethane
Trx	thioredoxin
UV	ultraviolet

Nomenclature

I_O	open pore current
I_B	current through pore partially blocked by an analyte
I_{RES}%	residual current as percentage of I_O: $I_{RES}\% = I_B / I_O \times 100\%$
$I_{RES}^{N-, L2}$%	I_{RES}% of the current blockade of level 2 for N terminus-first translocation (in the experiment of co-translocational protein unfolding through nanopore)
$I_{RES}^{C-, L2}$%	I_{RES}% of the current blockade of level 2 for C terminus-first translocation
$I_{RES}^{N-, L5}$%	I_{RES}% of the current blockade of level 5 for N terminus-first translocation
$I_{RES}^{C-, L5}$%	I_{RES}% of the current blockade of level 5 for C terminus-first translocation
ΔI_{RES}%	difference in I_{RES}% between current blockades

$\Delta I_{RES}^{L2}\%$	difference in $I_{RES}\%$ of the current blockade of level 2 between N and C terminus-first translocation: $\Delta I_{RES}^{L2\%} = I_{RES}^{N-, L2}\% - I_{RES}^{C-, L2}\%$
$\Delta I_{RES}^{L5}\%$	difference in $I_{RES}\%$ of the current blockade of level 5 between N and C terminus-first translocation: $\Delta I_{RES}^{L5\%} = I_{RES}^{N-, L5}\% - I_{RES}^{C-, L5}\%$
$I_{RES}^{oligo(dA)}\%$	$I_{RES}\%$ of the current blockade produced by oligo(dA)
$I_{RES}^{oligo(dC)}\%$	$I_{RES}\%$ of the current blockade produced by oligo(dC)
$\Delta I_{RES}^{oligo}\%$	difference in $I_{RES}\%$ of the current blockade produced by oligo(dA) and oligo(dC): $\Delta I_{RES}\% = I_{RES}^{oligo(dA)}\% - I_{RES}^{oligo(dC)}\%$
I_n	current noise (standard deviation of a Gaussian fit to all-points histogram for a current)
$I_n^{N-, L5}$	I_n of the current blockade of level 5 for N terminus-first translocation
$I_n^{C-, L5}$	I_n of the current blockade of level 5 for C terminus-first translocation
ΔI_n^{L5}	difference in I_n of the current blockade of level 5 between N and C terminus-first translocation: $\Delta I_n^{L5} = I_n^{N-, L5} - I_n^{C-, L5}$
f	normalized frequency of translocation events in s^{-1} μM^{-1}
t_{ON}	interevent interval (time between translocation events)
t_D	dwell time, time taken for translocation of an individual analyte
t_P	most probable dwell time, taken from the peak of a dwell time distribution

$t_P^{L2, N}$	t_P of the current blockade of level 2 for co-translocational unfolding of a native protein
$t_P^{L3, N}$	t_P of the current blockade of level 3 for co-translocational unfolding of a native protein
$t_P^{L4, N}$	t_P of the current blockade of level 4 for co-translocational unfolding of a native protein
$t_P^{L2, F}$	t_P of the current blockade of level 2 for co-translocational unfolding of a refolded protein in fully folded conformation
$t_P^{L2, I}$	t_P of the current blockade of level 2 for co-translocational unfolding of a refolded protein in partially folded conformation (intermediate)
$t_P^{L3, F}$	t_P of the current blockade of level 3 for co-translocational unfolding of a refolded protein in fully folded conformation
$t_P^{L3, I}$	t_P of the current blockade of level 3 for co-translocational unfolding of a refolded protein in partially folded conformation (intermediate)
$t_P^{L4, R}$	t_P of the current blockade of level 4 for co-translocational unfolding of a refolded protein
$k_{U\rightarrow I}^{L2}$	rate of formation of a protein intermediate evaluated according to the current blockade of level 2
$k_{U\rightarrow F}^{L2}$	protein folding rate evaluated according to the current blockade of level 2
$k_{U\rightarrow F}^{L3}$	protein folding rate evaluated according to the current blockade of level 3

CHAPTER 1

General introduction

The advances of experimental techniques to detect and analyze single molecules in solution have provided crucial insights into the molecular-scale processes. Nanopores, pores with the nanoscale inside diameter, can be used for the stochastic detection of a variety of analytes, generating information at the single-molecule level [1-5]. The analytes include metal ions, small organic molecules, nucleic acids and proteins. Therefore, nanopore stochastic sensing has emerged as an important technique playing a significant analytical role in disease diagnosis, drug screening, food processing and environmental monitoring. This chapter presents the background of principles and techniques underpinning nanopore stochastic

sensing and its applications, with an emphasis on the analysis of nucleic acids and proteins.

1.1 Nanopore stochastic sensing

Nanopore stochastic sensing is a new single-molecule technique with versatile applications in biophysics and biotechnology. The analyte molecules can be probed one at a time through the pore by electrophysiological techniques. Here follows an overview of the principles and technical aspects of nanopore stochastic sensing.

1.1.1 Principles of nanopore stochastic sensing

Nanopore stochastic sensing is based on the working principle of Coulter counters [6], which is a well-developed device used to measure the size and concentration of cell and virus particles. Nanopore, with the diameter three orders of magnitude smaller, is capable of sensing objects at the single-molecule level.

In nanopore stochastic sensing, a single nanopore is embedded in a membrane separating into *cis* and *trans* compartments, each containing conductive electrolytes (Figure 1.1). In an electric field across the pore, the resulting ionic current flowing through the pore can be detected and recorded by electrophysiological techniques (e.g., patch-clamp amplifier). The analytes

consisting of charged molecule driven through the pore by the electric field (known as electrophoresis) can interact with or block the pore, thus modulating (usually reducing) the ionic current. The transient changes in the ionic current, manifested as blockades, can provide information of the analytes. Full details of the designed apparatus and the experimental procedures can be found in the Methods in Chapter 5, Section 5.3.

The identity of the analyte can be determined or characterized by the dwell time of translocation (t_D) and the residual current level during the translocation (I_{RES}%), wherein the current blockade level (I_B) is expressed as a percentage of the open-pore level (I_O): $I_{RES}\% = I_B / I_O \times 100\%$. The concentration can be measured by the frequency of the blockade events (f), which is the inverse of the mean interevent interval (t_{ON}): $f = 1 / \bar{t}_{ON}$, when $\bar{t}_{ON} \gg \bar{t}_D$.

This technique offers nanometer-scale spatial resolution and microsecond-scale temporal resolution, thereby providing crucial insights into the molecular conformation and dynamics of the analytes.

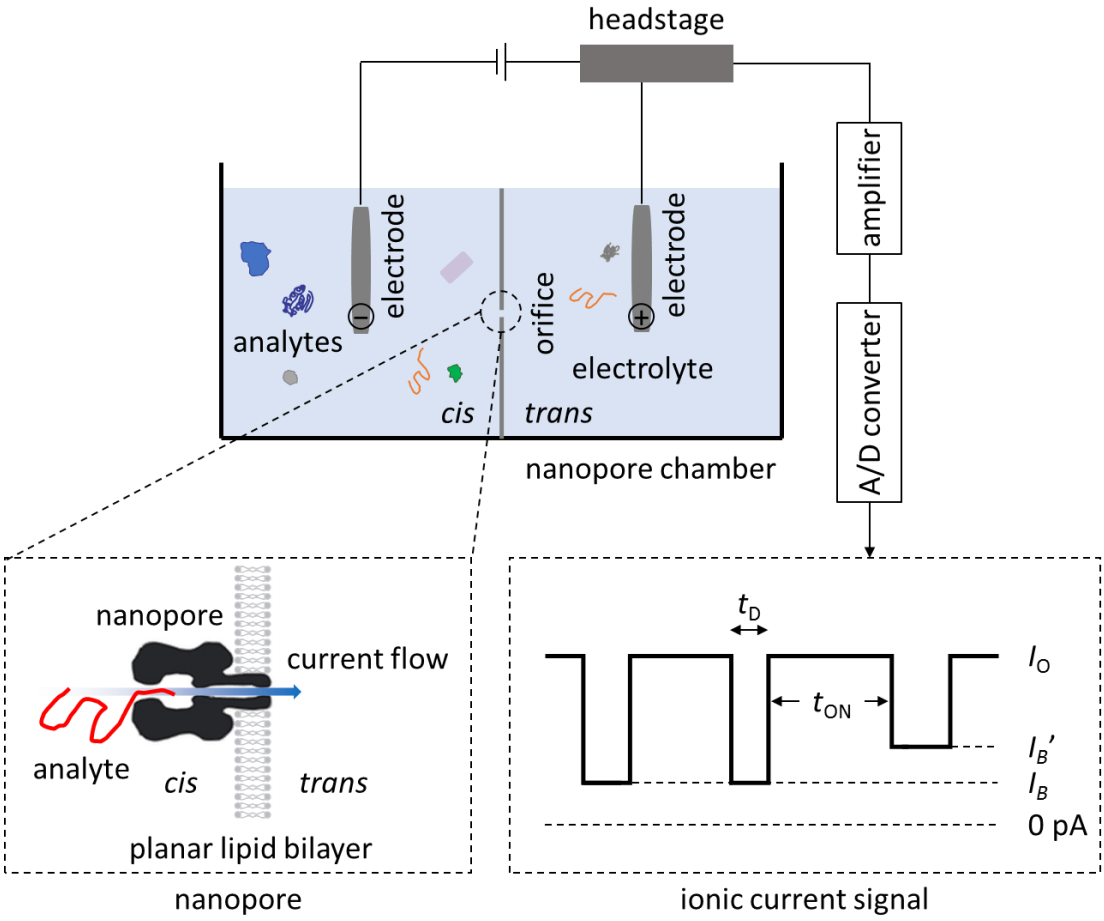


Figure 1.1 I Principles and experimental configuration of nanopore stochastic sensing. An orifice in the chamber which is used to support the planar lipid bilayer is inserted by a biological nanopore (magnified and shown as a left inset). The ionic current is measured by a patch-clamp amplifier and then converted into digital signals by an analogue-to-digital convertor (A/D convertor). The current signal for analytes translocating through the pore shows the open-pore level (I_O), the translocation duration (t_D), the current blockade levels (I_B) and the interevent intervals (t_{ON}) for individual translocation events (right inset).

1.1.2 Nanopores

The nanopore can be a naturally occurring biological pore inserted into a planar lipid bilayer [7], or an artificial pore fabricated in a solid-state membrane [8,9]. The commonly used biological nanopores include *Staphylococcal aureus* α-hemolysin (αHL), *Mycobacterium smegmatis* porin A (MspA), *Aeromonas hydrophila* aerolysin and *Escherichia coli* Curlin sigma S-dependent growth subunit G (CsgG) [10,11]. They are harvested in cells in large quantities, so that these pores are atomically consistent and readily engineered by site-directed mutagenesis [12], and they can be obtained at low cost. However, their lipid matrixes are mechanically fragile. Solid-state nanopores can be fabricated from silicon oxide, silicon nitride, aluminum oxide, graphene and polymer [13,14]. They have much higher robustness and durability but less reproducibility than the biological nanopores. Hybrid nanopores, coupling a biological pore [15,16] or a DNA origami pore [17,18] into a solid-state substrate, have also been designed to overcome the disadvantages for both types of nanopores.

The numerous applications of the engineered biological [1,10,11,19] and the fabricated solid-state nanopores [13,19-22] for stochastic sensing have been reviewed in the past. This thesis is focused on the analysis of nucleic acids and proteins using the αHL pore; unless otherwise stated, the nanopore in the following contents refers to the αHL pore.

1.1.3 α-hemolysin pore

The αHL is a mushroom-shaped transmembrane pore formed by self-assembly of seven identical 293-residue polypeptides [23]. The heptameric αHL pore consists of a cap domain, a stem domain and seven rim domains. The cap domain is positioned outside the membrane bilayer and participates extensive inter-subunit interactions (Figure 1.2 a, b), which are responsible for the robustness of the heptamer. The stem domain is a 14-strand antiparallel β-barrel and is spanning across the membrane (Figure 1.2 b). The rim domains protrude from the underside of the heptamer (Figure 1.2 c); they contain only a few inter-subunit interactions but are in close proximity with the membrane.

The dimensions of the αHL pore have been determined by crystallography to 1.9 Å resolution (Figure 1.2 a, b) [24]. The external dimensions of the pore are ~100 × 100 × 100 $Å^3$. The inner diameters of the pore vary with ~28 Å at the *cis* entrance, ~46 Å in the vestibule and ~26 Å in the β-barrel. The pore is narrowest at the junction of the vestibule and the β-barrel, with ~15 Å at the inner constriction.

The αHL pore is barely large enough to allow the passage of single-stranded nucleic acids and unfolded polypeptide chains, but small enough to block the double-stranded [25,26] and most proteins in folded state [27]. The inner diameters of the pore and the sizes of these translocating molecules are of the same order, such that the molecules can strongly interact with the pore, which is a key feat of nanopore sensing.

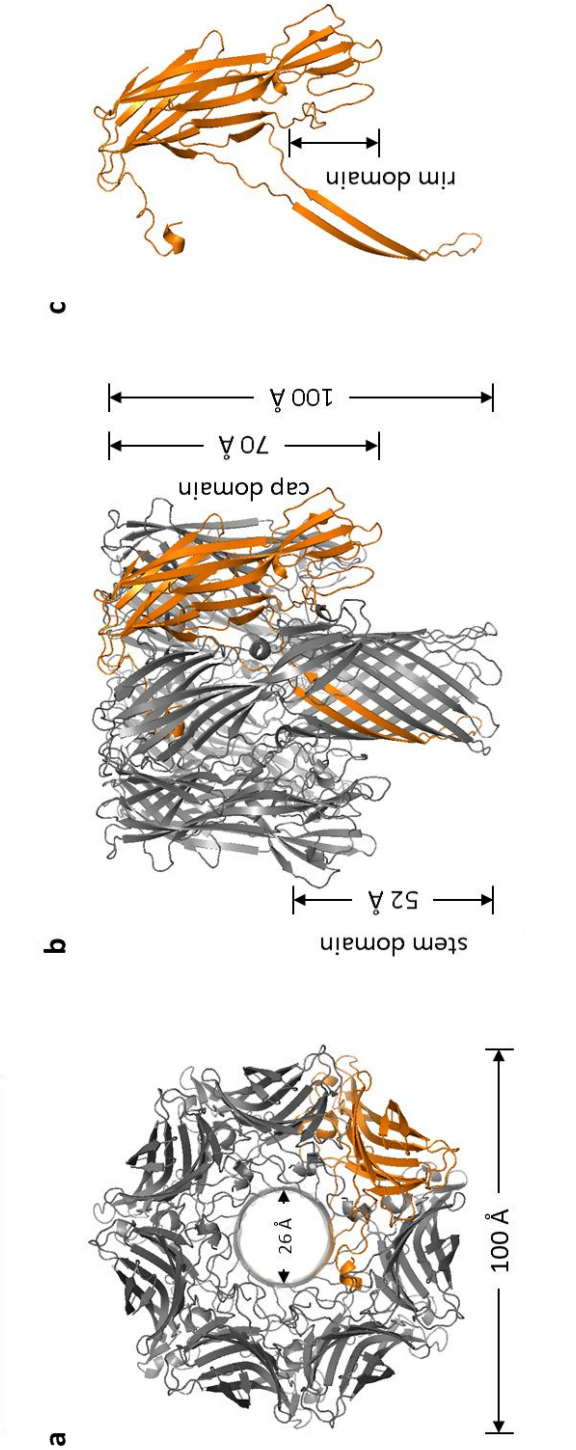


Figure 1.2 I Ribbon representations of the αHL pore. (a-b) The heptameric αHL pore showing the positions of the cap and the stem domains. (a) The view from the top of the structure and parallel to the sevenfold axis. (b) The view perpendicular to the sevenfold axis and parallel to the putative membrane bilayer. (c) Single subunit of the heptameric αHL pore showing the position of the rim domain. This view is from the "outside" of the heptamer. Taken from PDB structure code 7AHL and visualized using PyMOL.

1.2 Analysis of nucleic acids

Nanopores have served as versatile single-molecule sensors; the most compelling application is the analysis of nucleic acids and towards DNA sequencing. This elegant technology is based on a simple concept [25,28,29] that negatively charged single-stranded DNA (ssDNA) can be electrophoretically driven through the nanopore and the translocating nucleotides modulate the ionic current flowing through the pore. All four canonical nucleotides (adenine (A), cytosine (C), guanine (G) or thymine (T)), each with a particular molecular structure and size, are expected to be readily identified by the extent to which they reduce the ionic current.

The success of such measurements, however, relies extremely on attaining single-base resolution in DNA strands, and this proved to be a difficult challenge [30-34]. Here I cover the principles of nanopore detection of nucleic acid polymer translocation and the key advances that have shown promise for DNA sequencing.

1.2.1 Detection of polynucleotides translocation

Early proof-of principle experiments demonstrated that homopolymeric RNA and ssDNA can be readily discriminated using the wild-type αHL pore based on the residual currents and the translocation dwell times [25,35,36]. For example, when the homopolymers freely translocate through the pore, poly(rA) gives a higher residual current value and a longer translocation dwell

time than poly(rC) [35]; poly(dA) shows similar extent of current blockade but longer event duration compared with poly(dC) (Figure 1.3) [36,37]. Homopolymeric segments transition within a molecule can also be detected: in an experiment of poly(rA)$_{30}$(rC)$_{70}$ translocation, current blockades with two levels were observed, which reflected the poly(rA) to poly(rC) transition [35]. In addition, the current blockade contains information about the polynucleotide orientation (5' or 3' end-first) during the translocation [38-40].

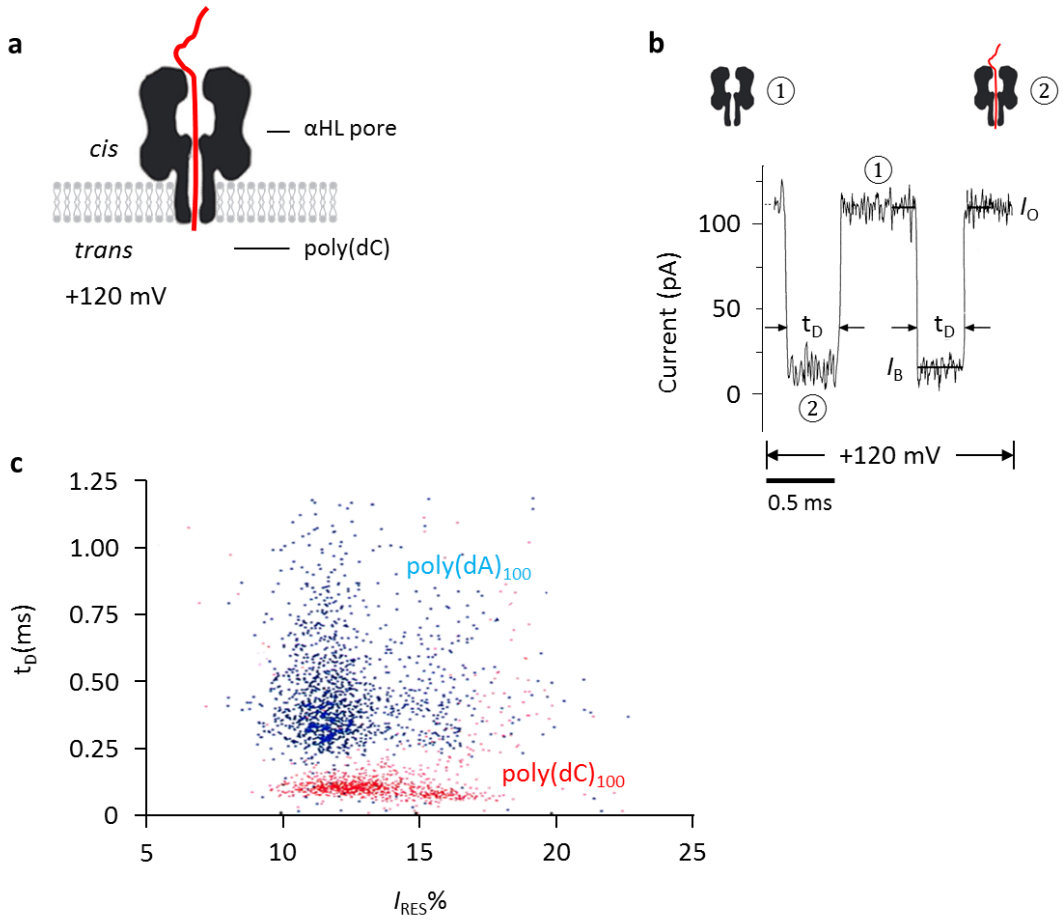


Figure 1.3 I Polynucleotides translocation through the αHL pore. (a) Schematics depicting the translocation of poly(dC) through the αHL pore under an applied voltage. (b) Upper panel: open-pore state and the translocation of the polynucleotide through the pore.

Lower panel: representative ionic current signal corresponding to two translocation events. The current blockade levels were observed during the translocation of polynucleotides through the pore (②). The current was restored to its original value when the polynucleotide left through the pore (①). (c) Scatter plots showing t_D versus I_{RES}% (I_{RES}% = I_B / I_O × 100%) for poly(dA)$_{100}$ (blue) and poly(dC)$_{100}$ (red). Reproduced with permission from Meller *et al*. [37].

For single-base identification, however, the polynucleotide translocation process is at remarkably high rates of several microseconds per base under typical experimental conditions (e.g. 1–22 μs base^{-1} at +120 mV and 25 °C) [37]; these rates are two to three orders of magnitude faster than the detection limits, on the order of milliseconds, of conventional electrophysiological techniques. Therefore, to resolve single bases with the ionic current detection system, the main challenge that has to be overcome is slowing polynucleotide translocation rate and controlling its motion through the pore [30,41]. The translocation rate could be reduced by using passive approaches, including decreasing the temperature [37], reducing the applied voltage [36,42] increasing the solvent viscosity [43], adding organic salts [44] or introducing positively charged residues within the pore [45,46]. However, these attempts at slowing the translocation rate are at the expense of reduced measured signal. Moreover, labelling of nucleotides with bulky groups enables the detection of multiple separate bases, but it remains challenging to label contiguous nucleotides in the polynucleotide strands [47]. Active approaches of using processive enzymes to regulate the translocation are discussed in Section 1.2.4.

1.2.2 Identification of mononucleotides by the nanopore equipped with a molecular adapter

Aside from adjusting the exterior parameters and nucleotide labelling, equipping the nanopore with molecular adapters is another passive method that can halt the rapid slip of the analyte molecules and enhance the detection resolution. It has been shown that unlabeled nucleoside monophosphates (2′-deoxyadenosine 5′-monophosphate (dAMP), 2′-deoxycytidine 5′-monophosphate (dCMP), 2′-deoxyguanosine 5′-monophosphate (dGMP), and 2′-deoxythymidine 5′-monophosphate (dTMP)) can be accurately identified by a modified αHL pore, either noncovalently fitted with an aminocyclodextrin [48] or covalently attached a cyclodextrin [49] within the lumen of the pore. These adapters constrict the pore and produce a substantial but incomplete pore block. The translocating mononucleotides, which can bind within the hydrophobic interiors of the adapters, augment the blocking extent, thereby enabling superior single-base identification.

The success in the identification of mononucleotides is one of the key feats for nanopore exonuclease sequencing system. In this system, the entrance of the nanopore is coupled to a processive exonuclease, which sequentially digests individual mononucleotides from the end of the DNA strand. The exonuclease-released “bases” are electrophoretically translocated through and read by the pore [50]. However, the problem of assuring all the cleaved mononucleotides are captured and pass through the pore in the correct order, without overtaking, limits the application of this system [30,51].

1.2.3 Single-base discrimination in stationary polynucleotides

In addition to the identification of the isolated mononucleotides during their translocation through the pore, polynucleotides have also been interrogated when they statically captured within the pore. Although immobilization would not allow the polynucleotides passing through the pore and therefore precludes the strand-sequencing, it allows detailed exploration of the feasibility of single-base resolution by nanopore. Immobilization of the analyte enables indefinite extension of the measurement time. The noise in the signal can be reduced by stringent filtration and, thereby, higher signal-to-noise ratio can be attended. The improved signal resolution is expected to allow the discrimination of small differences in the current blockades associated with different polynucleotides. Moreover, the polynucleotide strands are more stretched when immobilized within the pore in an electric field, and this limits their conformational flexibility affecting the signal.

The αHL pore has proven to be capable of discriminating immobilized homo-polynucleotides. In the streptavidin immobilization experiments [52-54], biotinylated polynucleotides have been attached to their 3’ ends by a protein, streptavidin, of larger size than the αHL pore inner diameter (Figure 1.4 a). In an electric field across the pore, the free 5’ end threads through the pore until the anchor halts translocation, and thus the polynucleotide can be statically captured within the pore. The stretched polynucleotide partially blocks the flow of current, leading to a sequence-dependent residual current. For the wild-type αHL pore, poly(dA) blocks the pore to a lesser extent than poly(dC)

(Figure 1.4 b, d) [52]. Whereas, an engineered αHL pore (E111N/K147N) gives the opposite blocking results (Figure 1.4 c, e) [53]. With the elimination of the electrostatic interactions and increased space at the central constriction of the pore, it attains a significantly larger residue current difference between deoxypurine and deoxypyrimidine polymers. Moreover, both wild-type [54] and E111N/K147N αHL pore [53] are able to resolve single base substitutions in the homopolymeric background.

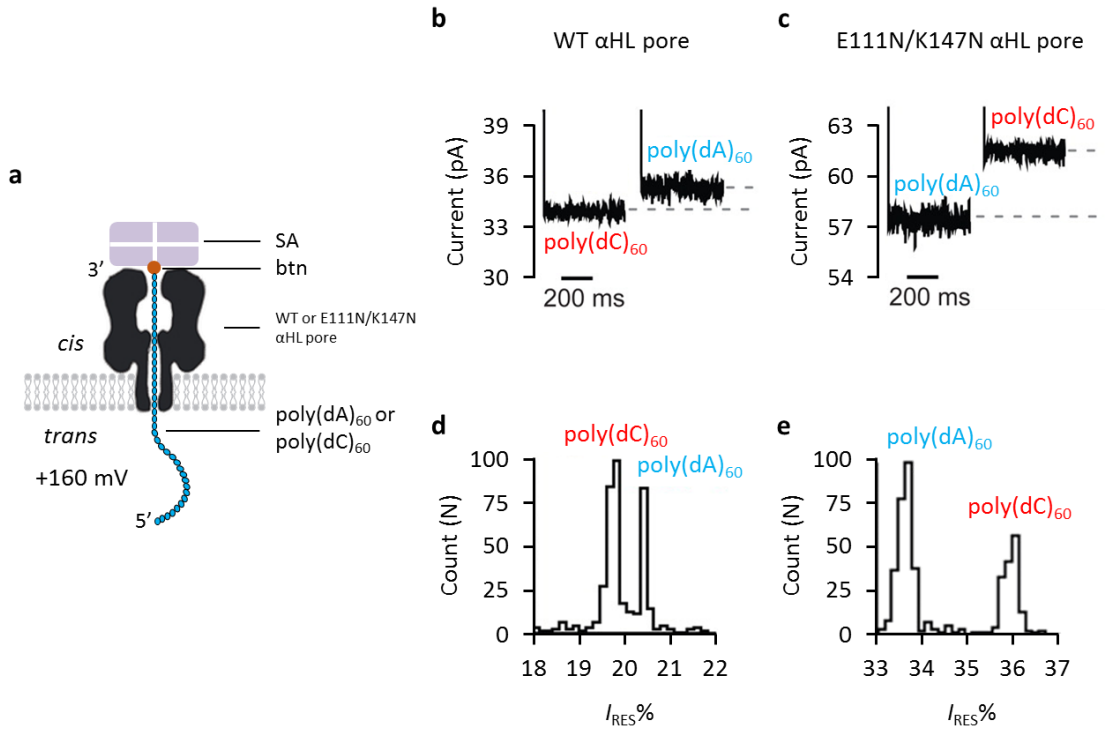


Figure 1.4 I Discrimination of immobilized homo-polynucleotides by the αHL pore. (a) Schematics depicting the homopolymer ($poly(dA)_{60}$ or $poly(dC)_{60}$) immobilized within the pore through the use of a biotin-streptavidin linkage (only the first 40 nucleotides of the 60-nucleotide-long sequence are shown). (b-c) Residual current levels produced by $poly(dA)_{60}$ and $poly(dC)_{60}$ when immobilized within the wild-type and E111N/K147N αHL pores. (d-e) Event histogram of I_{RES}% values produced by $poly(dA)_{60}$ and $poly(dC)_{60}$ when immobilized within the wild-type and E111N/K147N αHL pores. SA: streptavidin, btn: biotin. Reproduced with permission from Stoddart *et al*. [53].

Analysis of heteropolymeric sequence adds more complexities. There are three sensing sites within the β-barrel of the wild-type αHL pore: near the constriction, in the middle of the β-barrel, and near the *trans* exit [55,56]; the β-barrel allows ~12 nucleotides occupied [57]. The current level of one blockade is contributed from all three sensing sites and, in each sensing site, individual base identification can be affected by the neighboring nucleotides. Nevertheless, with the E111N/K147N αHL pore and at a specific position within the stretched stationary polynucleotide, where the base meets one of the sensing sites (Figure 1.5 a), all four canonical bases were distinguished (Figure 1.5 b) [53]. Soon after, other engineered αHL pores have been exploited to further purposefully tune the recognition abilities, such as E111N/K147N/M113Y (with one of the sensing sites strengthen), E111N/K147N/M113G (with one of the sensing sites weaken) [55], and truncated-barrel mutants [58]. Modified bases, such as 5-methylcytosine and 5-hydroxymethylcytosine were readily identified by using these engineered pores [59].

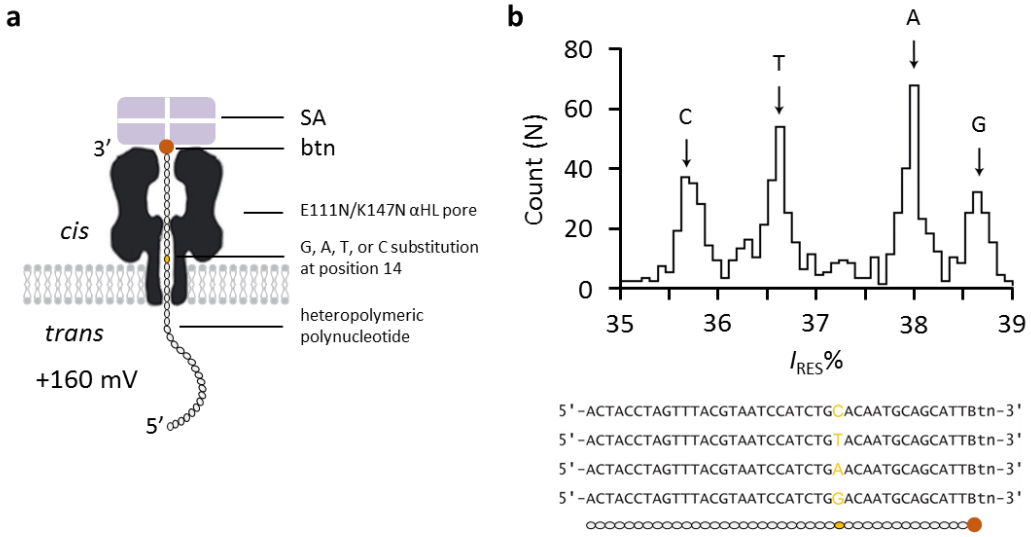


Figure 1.5 I Single-base discrimination in an immobilized hetero-polynucleotide by the E111N/K147N αHL pore. (a) Schematics depicting the heteropolymeric sequence immobilized within the pore through the use of a biotin-streptavidin linkage. (b) Event histogram of I_{RES}% values produced by four heteropolymeric sequences when immobilized within the E111N/K147N αHL pore. The bases C, T, A and G in the heteropolymeric sequence each gave a different I_{RES}% value. Reproduced with permission from Stoddart *et al*. [53].

In other work, polynucleotides have been immobilized within the pore by using terminal hairpins [39,60,61]. However, multiple dynamic blockade levels were produced, because the hairpin could interfere with the pore [60]. Another engineered nanopore, MspA pore, with a single and shorter (60 Å) sensing site [62], afforded superior recognition among all four different bases within the polynucleotide chain [63,64].

These encouraging results suggest a promise of nanopore strand-sequencing system. The stationary polynucleotide electrophoretically stretched within the pore is, in some respects, in a state analogous to the ratcheting movement controlled by a processive enzyme, in which each ratcheting step allows

sufficient delay time for single-base resolution.

1.2.4 Ratcheting of polynucleotide through the nanopore

Upon achieving the single-base discrimination, controlling the movement of polynucleotide through the nanopore is another essential criterion for the reading of the bases within the polynucleotide. As was conceived in the first nanopore sequencing patent [28], DNA polymerases have proven to be ideal motors to substantially slow the translocation rate and precisely control the ssDNA motion, one nucleotide at a time, through the pore.

In this system, the transport of DNA substrate into or out of the pore is regulated by an applied electric field and the polymerase activity. Initial attempts of using the *Escherichia coli* exonuclease I [65], Klenow fragment of *Escherichia coli* DNA polymerase [66] or bacteriophage T7 DNA polymerase [67] bound to the DNA substrates substantially reduced the ssDNA translocation rate by orders of magnitude. Further, single-base resolution was achieved with the DNA polymerase activity of nucleotide-by-nucleotide ratcheting the ssDNA through the pore [67-69]. However, these polymerases suffer from the low stability with the DNA substrates under an electric field. Subsequent advances overcome this limitation by using a bacteriophage Φ29 DNA polymerase [70,71] and a “blocking oligomer” strategy [72], in which ~500 DNA strands can be analyzed by individual nanopores.

To prevent the DNA synthesis in bulk phase, a blocking oligomer bearing two

acridine residues was hybridized adjacent to the DNA primer-template junction; to facilitate nanopore capture of polymerase-DNA complex in the electric field, the template was designed extended with a long 5′ overhang. A stretch of abasic residues, which could cause a residual current increase, was incorporated in the DNA substrate to assist the analysis of DNA displacement through the pore (Figure 1.6 a) [73]. Upon electrophoretic capture of the polymerase-DNA complex, the blocking oligomer was mechanically unzipped as the template moving forward through the pore (Figure 1.6 b). The removal of the blocking oligomer allowed polymerase catalyzing template-directed DNA replication, which ratcheted the template back through the pore. A wide variety of distinct current blockade levels were produced during the DNA template pass through the pore in both 3′-to-5′ voltage-driven unzipping and 5′-to-3′ replication-driven ratcheting direction (Figure 1.6 c). The translocation rate is reduced to ~10 ms base^{-1}, which is appropriate for single-base resolution. Also, forward and reverse ratcheting of the DNA strand through the pore provides double-pass identification of individual nucleotides. However, the recorded multiple current blockade levels are not directly associated with four nucleotides. Therefore, the sequence analysis of a given DNA requires sophisticated decipherment, which relies on the comparison with the known sequence in each case.

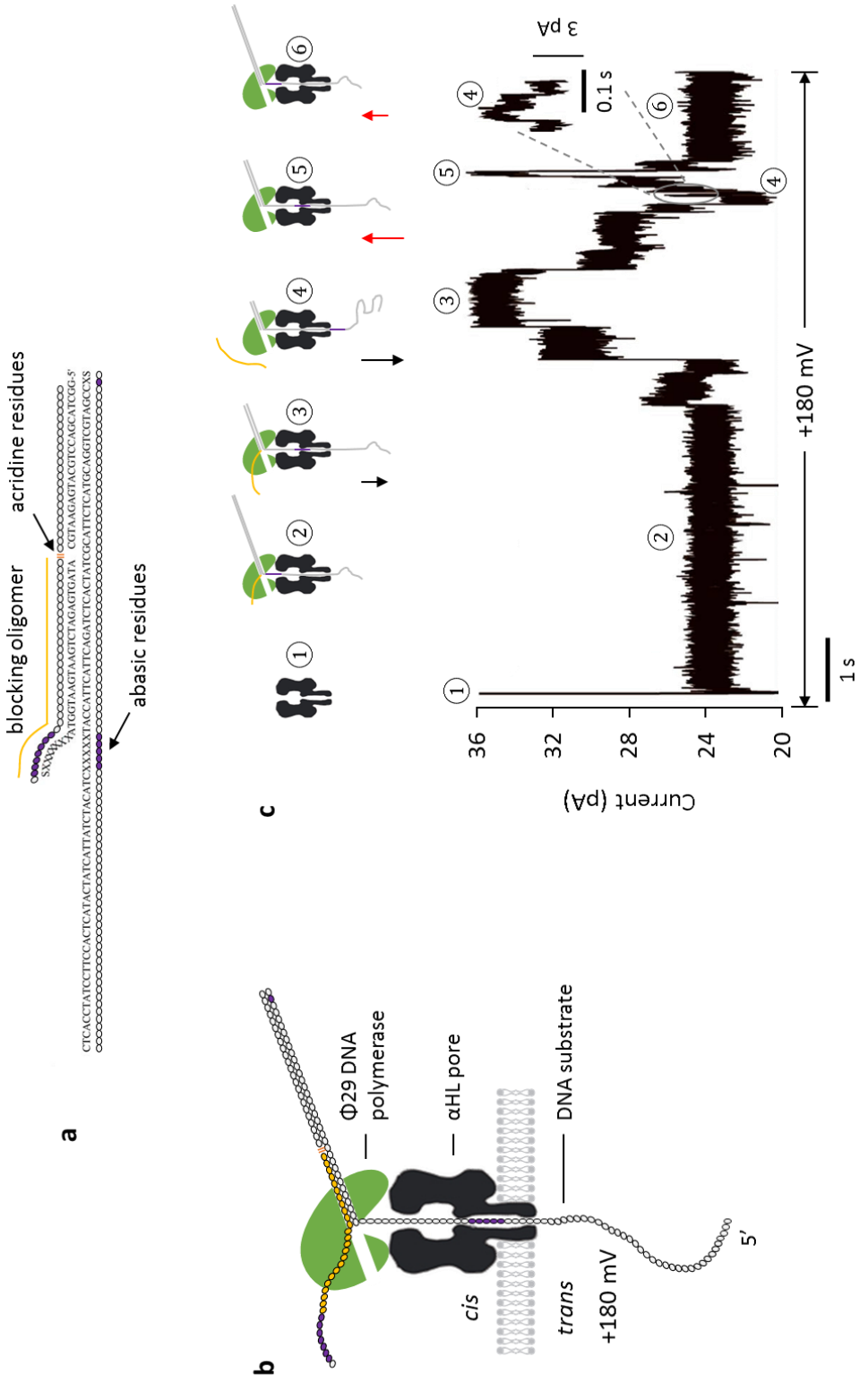

a
blocking oligomer
acridine residues
abasic residues
b
Φ29 DNA polymerase
αHL pore
DNA substrate
cis
trans
+180 mV
5′
c
Current (pA)
36
32
28
24
20
1 s
+180 mV
3 pA
0.1 s

Figure 1.6 I Forward and reverse ratcheting of DNA through the αHL pore. (a) The DNA substrate protected by a blocking oligomer. The blocking oligomer was capped by two acridine residues at its 5' end. The DNA template contains five abasic residues. (b) Schematics depicting the ratcheting of ssDNA through the αHL pore, controlled by the Φ29 DNA polymerase. (c) Upper panel: forward and reverse ratcheting of DNA through the pore. Lower panel: representative ionic current signal corresponding to the ratcheting process. ①–⑥ correspond to each other in both panels. Reproduced with permission from Cherf *et al*. [72].

Further studies using the same strategy but with an engineered MspA pore achieved a higher single-base resolution [74-78]. The success of using DNA polymerases to ratchet DNA strands through the pore, attaining slow translocation speed in single-nucleotide steps, progresses a crucial step towards nanopore strand-sequencing [79].

Overall, the remarkable single-molecule sensitivity of nanopore holds the promise of continuous single-base resolution during the ssDNA ratcheting through the pore. Therefore, nanopore is emerging as one of the most attractive technologies for next-generation DNA sequencing [33,34,80], and it has been further pursued and commercialized by Oxford Nanopore Technologies [81,82].

1.3 Analysis of protein unfolding and translocation

While the nanopore technology has been routinely used in a variety of ways to analyze nucleic acids, its applications in the analysis of proteins has also emerged as a rapidly growing research area. These include detecting protein-nanopore and protein-protein interactions, probing protein and peptide translocation, analysis of protein conformations and, more recently, sensing co-translocational protein unfolding [83,84]. This section is only focused on protein unfolding and translocation through the nanopore.

1.3.1 Translocation of peptide chains

The sensing of peptide chains with nanopore adds molecular complexities compared to the ssDNA due to the variety of amino acid side chains and the wide range of chemical properties. This was first demonstrated in a study on the translocation of synthesized peptides through the αHL pore [85]. The peptides contain “Gly−Pro−Pro” sequence repeat(s), and exit as mixtures of single, double, or collagen-like triple helices. They blocked the pore to different extent and for different duration during their translocation in an electric field. Therefore, the conformation of individual peptides was identified based on the residual current and the dwell time of their ionic current signatures.

In a subsequent study, positively charged α-helical peptides were electrophoretically driven through the pore [86]. The event frequency was

heavily dependent on the peptide length and the applied voltage across the pore: an increase in event frequency as a result of higher voltage and a decrease for a longer chain. The translocation was slowed due to the electrostatic and steric interactions within the narrow β-barrel of the pore. An engineered αHL pore (K131D/K147D) introducing negatively charged residues at the entrance and exit of the β-barrel facilitated the passage of the polypeptides [87]. Moreover, the translocation of negatively charged α-helical peptides has also been investigated through the αHL pore and an aerolysin pore (a nanopore lacking the vestibule but with a longer stem of a similar diameter [88]) [89]. This work suggested that the vestibule of the αHL pore helps to orient the peptides for translocation.

For peptides of larger size than the pore lumen, unfolding and unzipping processes become key to the translocation. Peptides with β-hairpin structure gave a longer translocation duration than the peptides without the hairpin structure. The dwell time strongly depended on the applied voltage, which suggested that the unfolding of the hairpin could be a rate limiting step [90].

These findings have important implications for understanding the mechanisms of the important biological process of peptide transport and are essential to lay the groundwork for the analysis of protein translocation.

1.3.2 Translocation of unfolded proteins

While the αHL pore allows the translocation of unfolded peptide chains, its

constriction (diameter ~15Å) is smaller than most proteins in folded state, thereby forbidding the passage of native proteins. On the other hand, the pore has an ability of insusceptible to chemical and thermal denaturing conditions [26,91,92]. Therefore, the αHL pore has been viewed to be an ideal candidate for the use as a model system to track protein unfolding in the presence of chemical denaturing agent.

In one study, maltose-binding protein (MalE), a negatively charged large protein (overall dimensions of 30 × 40 × 65 ×$Å^3$) was used, and it was denatured by guanidium hydrochloride in gradient concentrations to achieve partially and fully unfolded conformations (Figure 1.6 a) [27]. When these proteins were electrophoretically driven through the pore, a series of current blockades with different duration were observed (Figure 1.6 b). The short blockades were generated by the transit of the fully unfolded proteins: the event frequency increased as the concentration of the denaturing agent increased, which produced a sigmoidal unfolding curve (Figure 1.6 c), but their dwell times remained constant. The long blockades represent the translocation of partially unfolded conformations: their dwell times decreased as the proteins were more unfolded facilitated by the denaturing agent. The translocation of protein from both *cis* and *trans* sides of the αHL pore, and through the aerolysin pore, yield the identical unfolding curves [93,94]. However, the unfolding curve of a destabilized variant MalE 219 is shifted toward the lower concentration of denaturing agent compared to the wild-type. Protein unfolding as a function of temperature has also been investigated and obtained a similar melting point as in the bulk experiment [95].

These results illustrate that the protein unfolding process, characterized by nanopore translocation, is not dependent on the entry side, geometry or net charge of the pore, but is determined by the stability of the protein analyte.

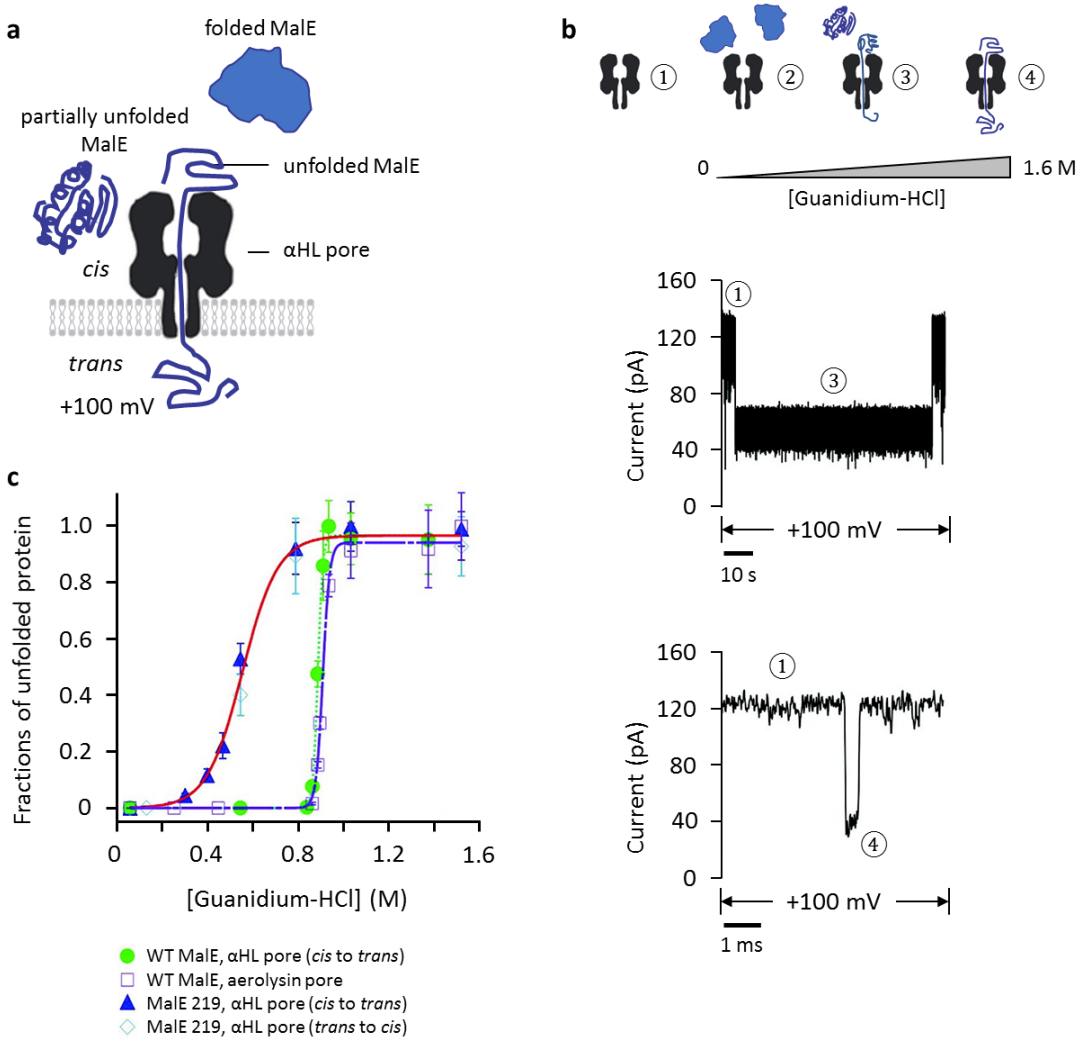


Figure 1.6 I Translocation of the unfolded proteins through the pores. (a) Schematics depicting the translocation of the unfolded MalE through the αHL pore. The driving force is the electric force. (b) Upper panel: detection of partially unfolded and fully unfolded MalE in the presence of guanidium hydrochloride. Lower panel: representative ionic current signals corresponding to the translocation of partially folded and unfolded MalE. ①–④ correspond to each other in both panels. (c) Unfolding fractions as a function of guanidium hydrochloride concentration. The curves for the wild-type MalE was obtained with the αHL pore and the aerolysin pore, and for the destabilized variant MalE 219 was obtained with the αHL pore (translocation from two sides of the pore). Reproduced with permission from Merstorf *et al*.

[94].

1.3.3 Unfoldase-assisted co-translocational protein unfolding

The translocations of proteins through the nanopore described in the preceding text were achieved by unfolding the proteins in the bulk phase bathing the pore. A crucial progress in the analysis of protein translocation has been made since 2013, which extends this technique to another level: nanopore-mediated protein unfolding and translocation [96,97]. Protein unfolding can occur by mechanically driven during its translocation through the pore, in a process different fundamentally from that accelerated by chemical denaturation or thermal perturbation.

As with the early exemplary work of enzymatic ratcheting of DNA through the pore (Section 1.2.4), enzyme-control is one strategy to pull the protein unfolding and translocation through the pore. For example, AAA+ enzyme ClpX, a bacterial unfoldase, can provide continuous mechanical force (~20 pN) at the expense of ATP hydrolysis to unfold protein substrates through its narrow central channel [98,99]. In one attempt, an ubiquitin-like protein, Smt3 domain, was attached with a polyanionic peptide tail and an ssrA peptide tag (ClpX targeting motif) at its C terminus [96]. In an applied positive voltage across the αHL pore, the negatively charged polymer was threaded through the β barrel of the pore and facilitated the capture and retention of the Smt3 into the *cis* entrance (Figure 1.7 a). At this step, the folded Smt3 was too large to traverse the pore, but the ssrA tag was able to proceed to the *trans* side

and was specifically bound to the ClpX. Next, in the presence of ATP, the unfoldase pulled the protein progressively unfolding, coincidently with the unidirectional translocation through the pore. A characterized ion current signature with four progressive levels was generated during this process (Figure 1.7 b): level 1 represents the open pore, level 2 represents the capture of Smt3, level 3 represents a preunfolding state mediated by ClpX, and level 4 represents the unfolded protein held within the pore. The current is restored (level 1) when the protein substrate leaves through the pore.

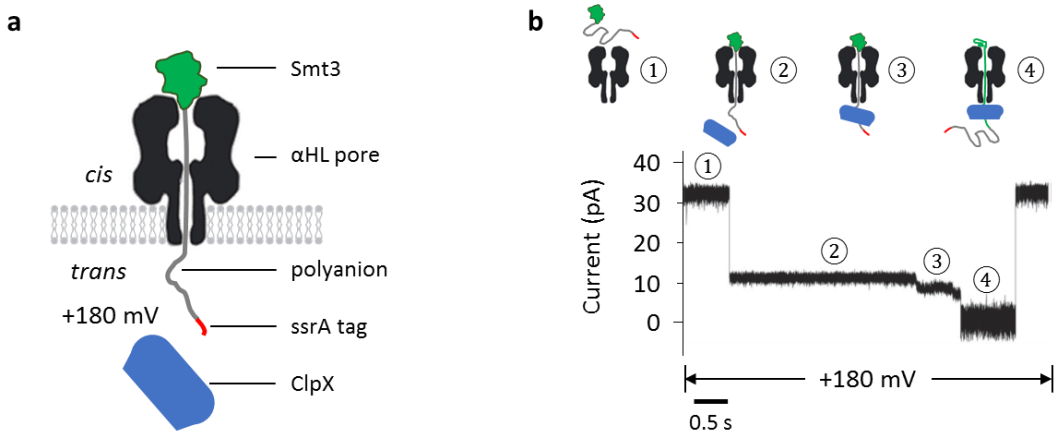


Figure 1.7 I Unfoldase-assisted co-translocational protein unfolding. (a) Schematics depicting the capture and retention of the Smt3 into the *cis* entrance of the αHL pore. (b) Upper panel: co-translocational unfolding of the Smt3 through the αHL pore, pulled by the ClpX. Lower panel: representative ionic current signal corresponding to the ClpX-assisted Smt3 translocation. ①–④ correspond to each other in both panels. Reproduced with permission from Nivala *et al*. [96].

This model system has been further developed to allow the identification of protein domains and the discrimination among protein variants (Figure 1.8 a) [100]. The unfoldase ClpX is supplemented coaxially with ClpP to form a more

robust ClpXP proteasome-like complex [98,99,101]. An engineered protein construct comprising four folded domains – two Smt3, one green fluorescent protein and one titin fragment – in series along strand was investigated. The recorded ionic current traces demonstrated distinctive patterns that reflected individual domains within the strand (Figure 1.8 b). In addition, sequence-dependent variations within those domains, such as point mutations, truncations and rearrangements could be classified at > 85% accuracy.

This unfoldase-assisted approach shows the ability of reproducibly pulling the designed protein unfolding and translocation through the pore, thereby achieving long reads of individual protein strands. However, in this study, protein analytes need to be genetically fused into a polyanion tail and an affinity tag, and it is likely to change the protein native structure.

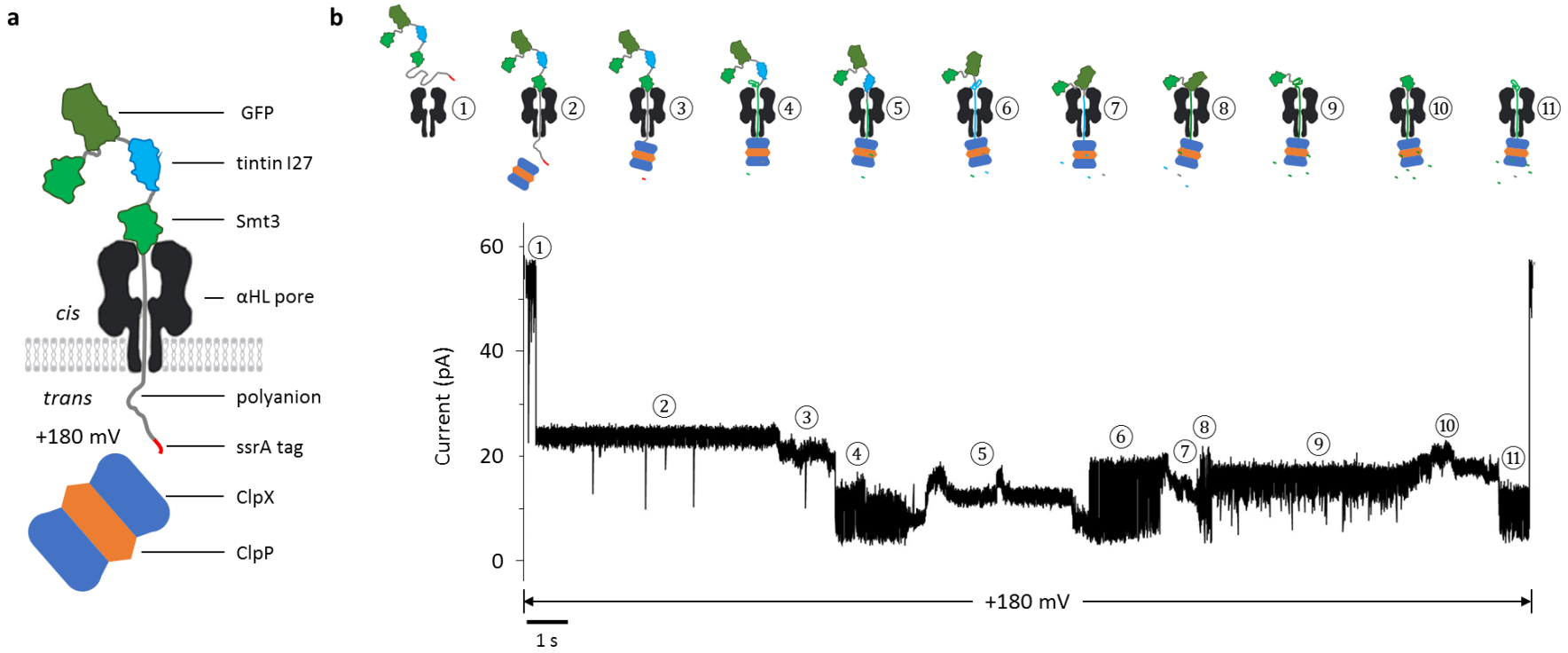


Figure 1.8 I Identification of protein domains during the translocation through the αHL pore. (a) Schematics depicting the capture and retention of the engineered protein construct into the *cis* entrance of the αHL pore. (b) Upper panel: translocation of the construct through the αHL pore, pulled by the ClpXP. Lower panel: representative ionic current signal corresponding to the ClpXP-assisted translocation of the protein construct. ①–⑪ correspond to each other in both panels. Individual domains within the construct were identified according to the ion current patterns. Reproduced with permission from Nivala *et al*. [100].

1.3.4 Potential-driven co-translocational protein unfolding

Potential-driven is another elegant approach to address the issue of protein translocation. In an electrical field across the pore, the unfolding and translocation of protein can be initiated by a charged leader sequence [102,103]. A negatively charged 30-mer oligonucleotide is capable of providing a tunable driving force, which is estimated at 5–10 pN at +140 mV [42,104,105], and it can be utilized to pull the protein unfolding and translocation through the pore.

The co-translocational unfolding of a model protein, thioredoxin, through the αHL pore has been investigated with this approach [97,106]. In the experiment, a thioredoxin stabilized variant containing an N- or C-terminal cysteine was used. A thiol-modified oligonucleotide is covalently attached to the cysteine at the protein terminus by disulphide bond formation. In an applied positive voltage, the oligonucleotide leader sequence experienced an electrical driving force and pulled the folded protein against the entrance of the pore, which presents as a physical barrier, thereby forcing the protein to unfold and traverse through the pore (Figure 1.9 a, c). The translocation is unidirectional, from either N or C terminus, depending on which terminus is attached to the oligonucleotide leader sequence.

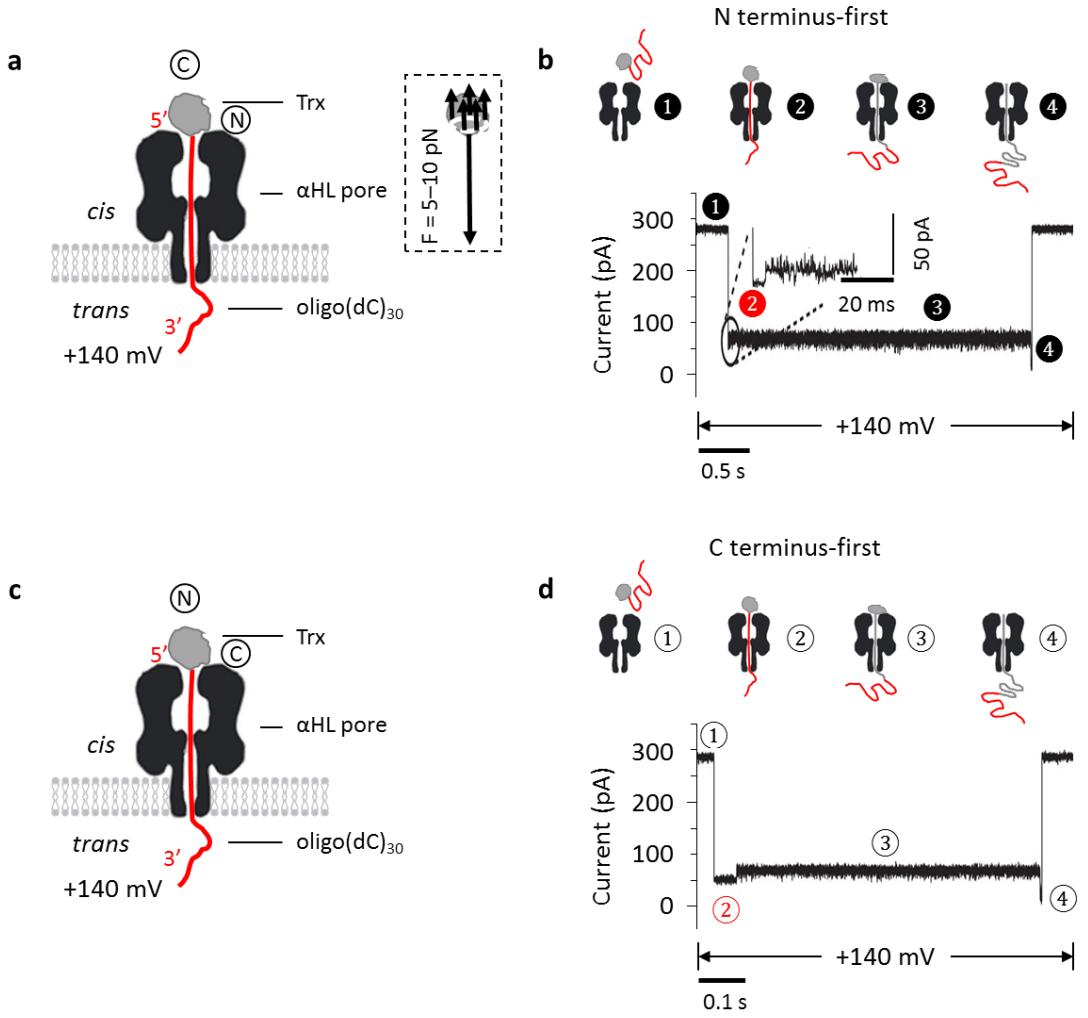


Figure 1.9 I Potential-driven protein unfolding and translocation through the αHL pore. (a) Schematics depicting N terminus threading of the oligonucleotide-thioredoxin conjugate into the *cis* entrance of the αHL pore. (b) Upper panel: co-translocational unfolding of the thioredoxin through the αHL pore, pulled by the oligonucleotide. Lower panel: representative ionic current signal corresponding to the potential-driven thioredoxin translocation. ①–④ correspond to each other in both panels. (c) Schematics depicting C terminus threading of the oligonucleotide-thioredoxin conjugate into the *cis* entrance of the αHL pore. (d) Upper panel: co-translocational unfolding of the thioredoxin through the αHL pore, pulled by the oligonucleotide. Lower panel: representative ionic current signal corresponding to the potential-driven thioredoxin translocation. ①–④ correspond to each other in both panels. Reproduced with permission from Rodriguez-Larrea and Bayley [97,106].

A series of distinct four-level ion current signatures were observed during the translocation of the oligonucleotide-protein conjugate (Figure 1.9 b, d). Each current blockade level has proved corresponding to a specific structural arrangement. Specifically, when the oligonucleotide was threaded into an unoccupied pore (level 1), it promoted the capture and retention of the folded protein sitting on top of the pore entrance (level 2). The force-bearing portion of the protein presented to the pore unfolded progressively, and the detached polypeptide chain moved within the pore (level 3). Lastly, the protein remaining segment outside the pore unfolded spontaneously and diffused through the pore (level 4).

The multi-step unfolding of thioredoxin during it translocation through the pore suggested the existence of intermediate(s), and it differs from those elucidated by denaturation experiments in bulk phase, for which two-state pathways have been demonstrated [107,108]. Further, the current blockade duration in level 2 is 20–30 times shorter when the protein is pulled from the N terminus as compared with the C terminus-first pulling. The level 3 for the N terminus-first, however, is bifurcated, with a short (approximate to C terminus-first) and a long dwell time (~100 times longer) distributions. These data indicate the different unfolding kinetics when the protein pulling from the N or C terminus and support a mechanism in which there is local unfolding near the terminus of the protein that enters the pore [109,110].

In the subsequent experiments, it has been further shown that protein multiple states of phosphorylation can be detected [111]. Phosphorylation produced a

decrease in the current noise during the protein translocates through the pore. Analysis of the residual current and the current noise enables the discrimination among unphosphorylated, monophosphorylated and diphosphorylated variants.

The potential-driven approach will be a fruitful area of research because it allows, with an optimized strategy of protein labelling, fishing the protein target from mixtures of biomolecules, and modifying the terminus with a charged leader sequence for nanopore translocation analysis, during which proteins' information (e.g. structure and modification) can be probed.

Taken together, the unique feature and promise of nanopore-based protein analysis, comparing with the commonly used mass spectrometry, allows probing unfragmented long protein strands at the single-molecule level during the translocation through the pore. Moreover, it enables studying protein unfolding induced by one-end pulling, which can be a process that differs from the global distortion by denaturing agents or the mechanical two-end pulling by force spectrometry. These findings suggest implications for understanding the mechanisms and evolution of protein unfolding and translocation through the biological channels in cells. More technically, these advances are one step towards the development of nanopore technology for protein sequence analysis and post-translational modification detection.

1.4 Objective

Nanopore-based sensing is attractive for DNA sequencing and, more recently, for protein analysis. In a series of elegant experiments, protein substrates can be unfolded as they are pulled through the pore with an oligonucleotide leader sequence. However, following the translocation, the protein is released into the solution and cannot be further examined to determine its refolded state.

The objective of current research is to determine the direction of protein translocation (N or C terminus-first) and to investigate protein refolding after the translocation. To achieve the directionality analysis, a model protein (thioredoxin) was attached at both N and C termini to different oligonucleotides. The direction of translocation can be ascertained with these distinguishable oligonucleotide tags. To capture the protein within the pore, we bound the two biotinylated oligonucleotides with traptavidin stoppers. This model system allows to investigate multiple cycles of protein unfolding, translocation and refolding in a controlled manner. This work aims to contribute to a better understanding of protein folding and unfolding *in vivo*.

CHAPTER 2

Bioconjugation of oligonucleotide to protein

2.1. Introduction

In order to enable the potential-driven unidirectional translocation of a protein through an αHL pore, an oligonucleotide must be attached to the terminus of protein (Chapter 1, Section 1.3.4) [97]. In an applied positive voltage across the pore, the negatively charged oligonucleotide facilitates the threading of the protein terminal end into the entrance of the pore, provides a driving force for protein unfolding and translocation, and prevents backward movement of the

protein-oligonucleotide conjugate. In the previous work, a 5'-thiol-modified oligonucleotide was activated with 2,2'-dipyridyl disulphide to yield 5'-*S*-thiopyridyl oligonucleotide, and this modified oligonucleotide was attached to the terminal cysteine, either N or C terminus, of a protein through a disulfide bond.

Taking advantage of the ability of the αHL pore to discriminate different oligonucleotides (Chapter 1, Section 1.2.3) [52-54], the translocation of protein pulled by the oligonucleotide leader sequence can be further developed to analyze the directionality of entry (N or C terminus-first). To achieve the discrimination, the protein is proposed to site-specifically attach to two different oligonucleotides at its N and C termini. Recent work demonstrated the development of a bioorthogonal click condensation reaction between 6-amino-2-cyanobenzothiazole (CBT) and N-terminal cysteine (Figure 2.1) [112,113]. This thiol-based reaction occurs in fireflies with high efficiency and biocompatibility [112,114]. The resulting thiazolidine ring structure is highly stable in biological conditions [114,115].

In this chapter, a CBT derivative for crosslinking has been rationally designed. We used this site-specific N-terminal cysteine condensation reaction, followed by the disulfide bond forming reaction, to conjugate two different oligonucleotides carrying thiol modifiers to both terminal cysteines of a model protein (a disulfide-free mutant of thioredoxin), one to the N terminus and the other to the C terminus.

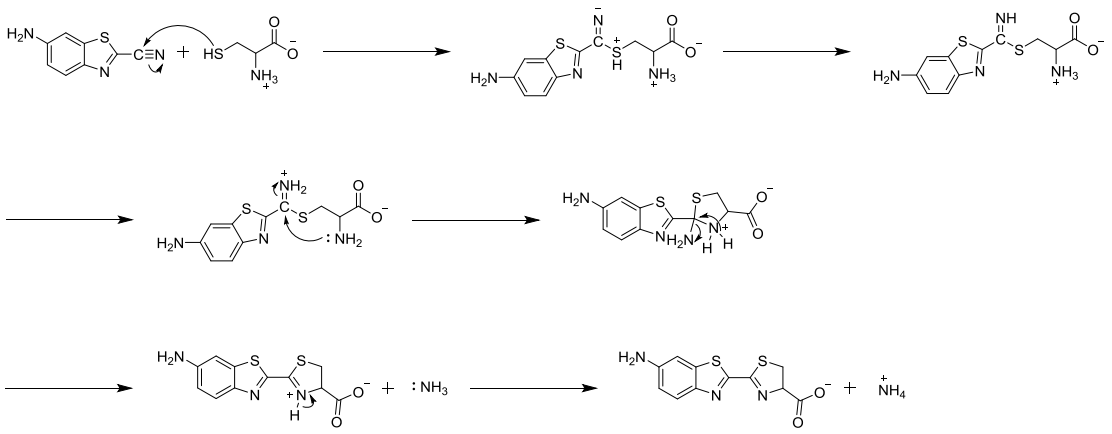


Figure 2.1 l Reaction mechanism between 6-amino-2-cyanobenzothiazole (CBT) and cysteine.

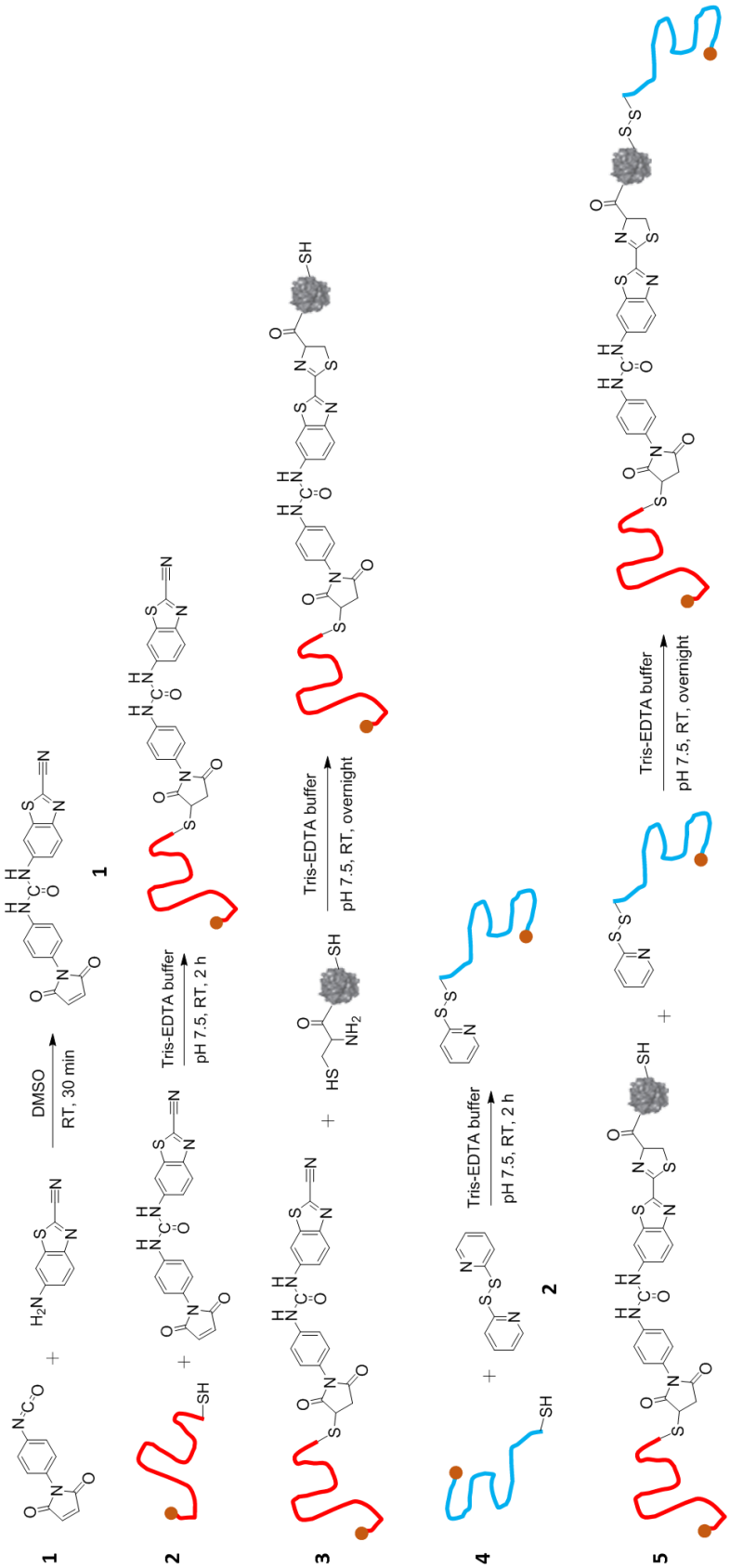


Figure 2.2 I Attachment of a different biotinylated oligonucleotide to each terminus of thioredoxin. step 1: preparation of the crosslinker, **1**; step 2: modification of btn-oligo(dC)$_{30}$-SH with **1**; step 3: bioconjugation of btn-oligo(dC)$_{30}$-**1** to the N terminus of thioredoxin to form btn-oligo(dC)$_{30}$-S1C-TrxV5-C109; step 4: activation of btn-oligo(dA)$_{30}$-SH with crosslinker, **2**; step 5: bioconjugation of btn-oligo(dA)$_{30}$-**2** to the C terminus of btn-oligo(dC)$_{30}$-S1C-TrxV5-C109 to form btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn. btn: biotin, shown as brown sphere; oligo: oligonucleotide, oligo(dC)$_{30}$ shown as red ribbon, oligo(dA)$_{30}$ shown as blue ribbon; S1C-TrxV5-C109: thioredoxin variant (S1C/C32S/C35S/A22P/I23V/P68A/C109), shown as grey mass.

2.2. Results

2.2.1 Synthesis of crosslinker

To apply this condensation reaction for the site-specific conjugation of a protein with an oligonucleotide, we chemically synthesized a new crosslinker, **1**, by coupling the amino group of CBT to the isocyanate group of PMPI (*p*-maleimidophenyl isocyanate) (Figure 2.2, step 1 and Figure 2.3). This novel crosslinker could be utilized to site-specifically conjugate a protein N-terminal cysteine to a thiol-modified oligonucleotide: the thiol on the oligonucleotide will react with the maleimide reactive moiety of **1** (Figure 2.2, step 2) and then the 1,2'-aminothiol on N-terminal cysteine protein will react with the cyano reactive moiety of **1** (Figure 2.2, step 3).

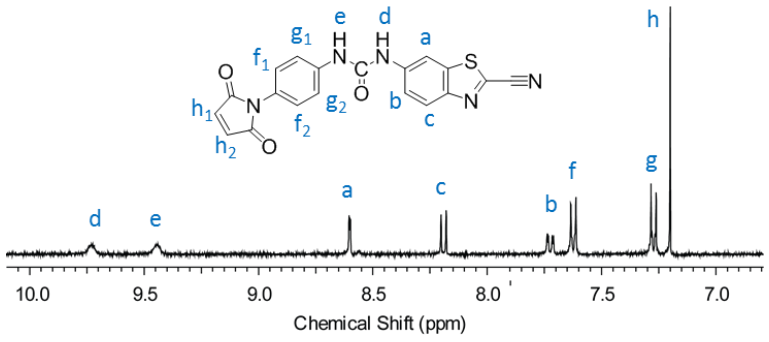


Figure 2.3 I ^{1}H NMR of the crosslinker, 1. ^{1}H NMR (400 MHz, DMSO-d_6): δ 7.20 (s, 2H), 7.26-7.28 (d, 2H, *J* = 8.85), 7.61-7.63 (d, 2H, *J* = 8.85), 7.72-7.74 (dd, 1H, *J* = 9.00, 1.98), 8.18-8.20 (d, 1H, *J* = 9.00), 8.60-8.61 (d, 1H, *J* = 1.98), 9.44(s, 1H), 9.73(s, 1H).

2.2.2 Bioconjugation of oligonucleotide to protein

A 5′-thiol-modified 3′-biotinylated oligonucleotide (Figure 2.4 a), btn-oligo$(dC)_{30}$-SH, was modified with **1** (Figure 2.2, step 2 and Figure 2.4 b-d) accordingly; another oligonucleotide, btn-oligo$(dA)_{30}$-SH, was activated with 2,2'-dipyridyl disulphide (crosslinker, **2**, Figure 2.2, step 4 and Figure 2.4 e-g). For the thioredoxin variant, we mutated the N-terminal residue to a cysteine and inserted a cysteine at the C terminus (S1C/C109). Two catalytic disulphide bonds were removed (C32S/C35S), which mimicked the reduced form of the protein. Three stabilizing mutations were also introduced (A22P/I23V/P68A) as described previously [108,116]. We referred to this variant as Trx V5-S1C/C109. The pyruvate adducts of the N-terminal cysteine, from bacterial expression, were removed with hydroxylamine to expose the 1,2'-aminothiol (Figure 2.5) [117].

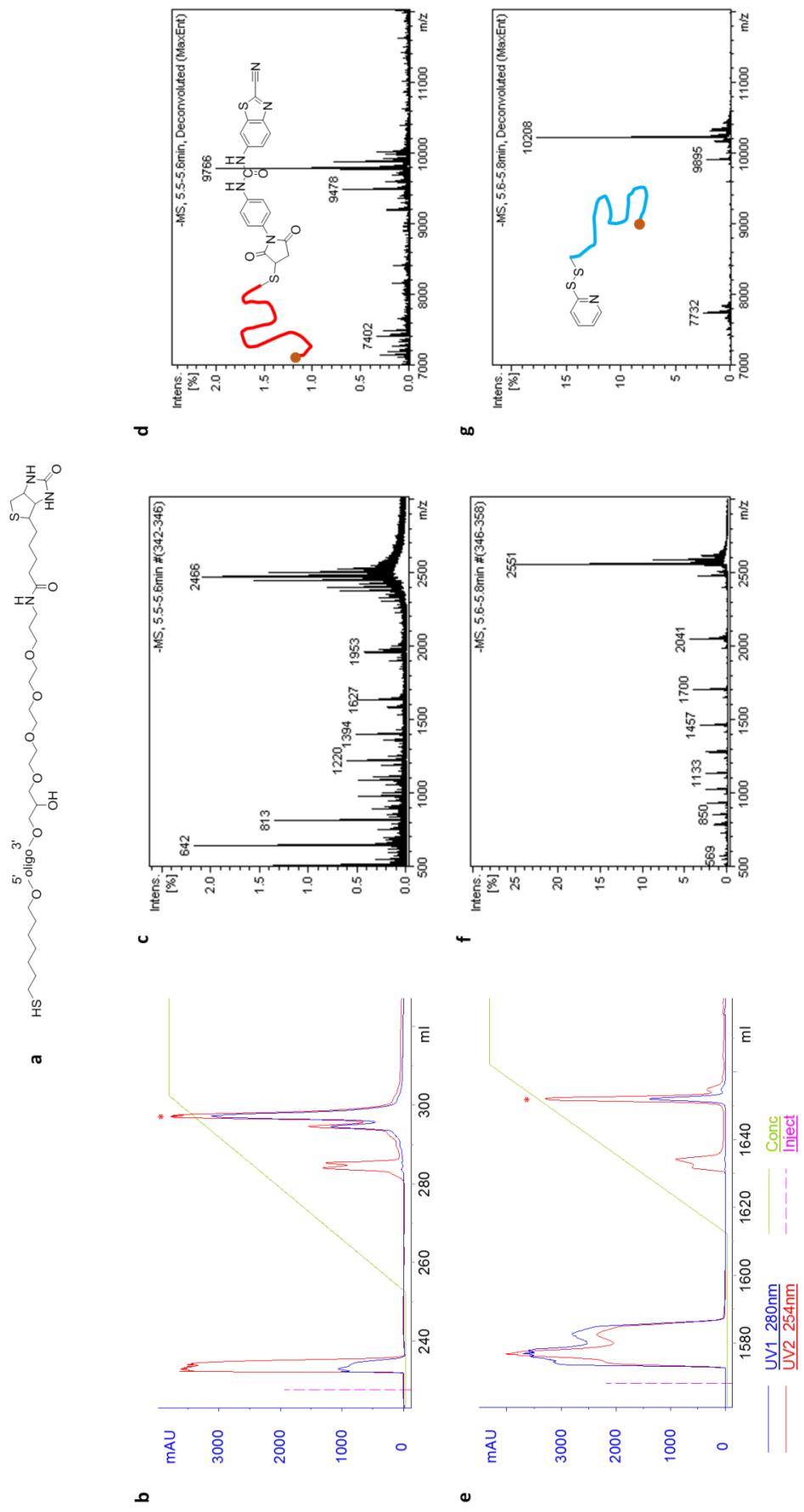

a
HS
5'
oligo
3'
OH
b
mAU
3000
2000
1000
0
240
260
280
300
ml
UV1 280nm
UV2 254nm
Conc
Inject
c
-MS, 5.5-5.6min #(342-346)
Intens.
[%]
642
813
1220
1394
1627
1953
2466
m/z
d
-MS, 5.5-5.6min, Deconvoluted (MaxEnt)
9766
9478
7402
e
f
-MS, 5.6-5.8min #(346-358)
569
850
1133
1457
1700
2041
2551
g
-MS, 5.6-5.8min, Deconvoluted (MaxEnt)
10208
9895
7732

Figure 2.4 I Oligonucleotide modification. (a) Chemical structure of the 5'-thiol (hexamethylene linker)- and 3'-biotin (triethylene glycol spacer)- modified oligo. The structure was produced by ChemDraw Professional 15.1. (b) Ion exchange chromatography purification and (c) ESI-MS and (d) of deconvoluted mass spectrum of btn-oligo$(dC)_{30}$-**1**. Calculated M = 9769, observed: 9766. (e) Ion exchange chromatography purification and (f) ESI-MS and (g) of deconvoluted mass spectrum of btn-oligo$(dA)_{30}$-**2**. Calculated M = 10209, observed: 10208. The desired product in each ion exchange chromatography is marked with a red asterisk.

CHAPTER 2

Bioconjugation of oligonucleotide to protein

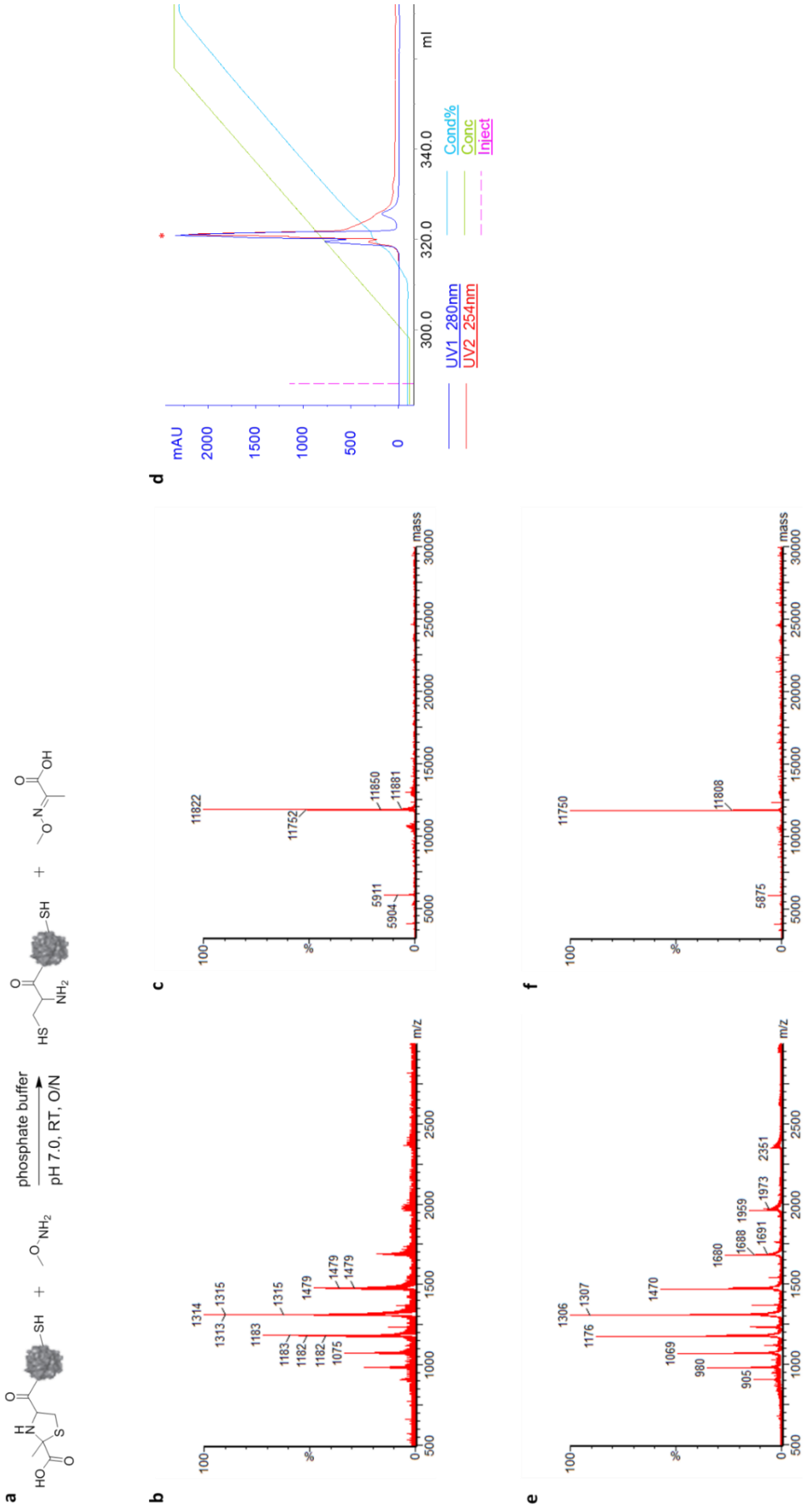


Figure 2.5 I Protein purification and thiazolidine deprotection. (a) Pyruvate-derived thioredoxin was deprotected with methoxyamine. (b) ESI-MS and (c) of deconvoluted mass spectrum of thioredoxin. Calculated $[M + H]^+$ = 11749. The observed mass of 11822 corresponds to the pyruvate adduct (+73 Da). (d) Ion exchange chromatography purification of thioredoxin. (e) ESI-MS and (f) of deconvoluted mass spectrum of thioredoxin deprotected with methoxyamine. observed: 11750. The desired product is marked with a red asterisk.

We then derivatized Trx V5-S1C/C109 at its N and C termini with btn-oligo$(dC)_{30}$-**1** and btn-oligo$(dA)_{30}$-**2** respectively. The btn-oligo$(dC)_{30}$-**1** was firstly linked to the N-terminal cysteine (S1C) of Trx V5-S1C/C109 via a cyanobenzothiazole condensation to obtain the mono-attached conjugate, btn-oligo$(dC)_{30}$-S1C-TrxV5-C109 (Figure 2.2, step 3 and Figure 2.6 a). Because the cyano group on **1** reacted site-specifically with 1,2-aminothiols on N-terminal cysteine, the C terminus (Cys-109) was not attached to btn-oligo$(dC)_{30}$-**1**.

Next, btn-oligo$(dA)_{30}$-**2** was linked to Cys-109 on the mono-attached conjugate, through a thiol-disulfide exchange reaction (Figure 2.2, step 5). The final construct (btn-oligo$(dC)_{30}$-S1C-TrxV5-C109-oligo$(dA)_{30}$-btn), was purified by ion exchange chromatography (Figure 2.6 b) and confirmed by SDS-PAGE gel and mass spectrometry (Figure 2.6 c, d). It should be noted that, in basic conditions, the maleimide linkage could become hydrolyzed at one of the two amides, giving a +18 peak in the mass, but leaving the adduct otherwise intact (Figure 2.6 e).

CHAPTER 2

Bioconjugation of oligonucleotide to protein

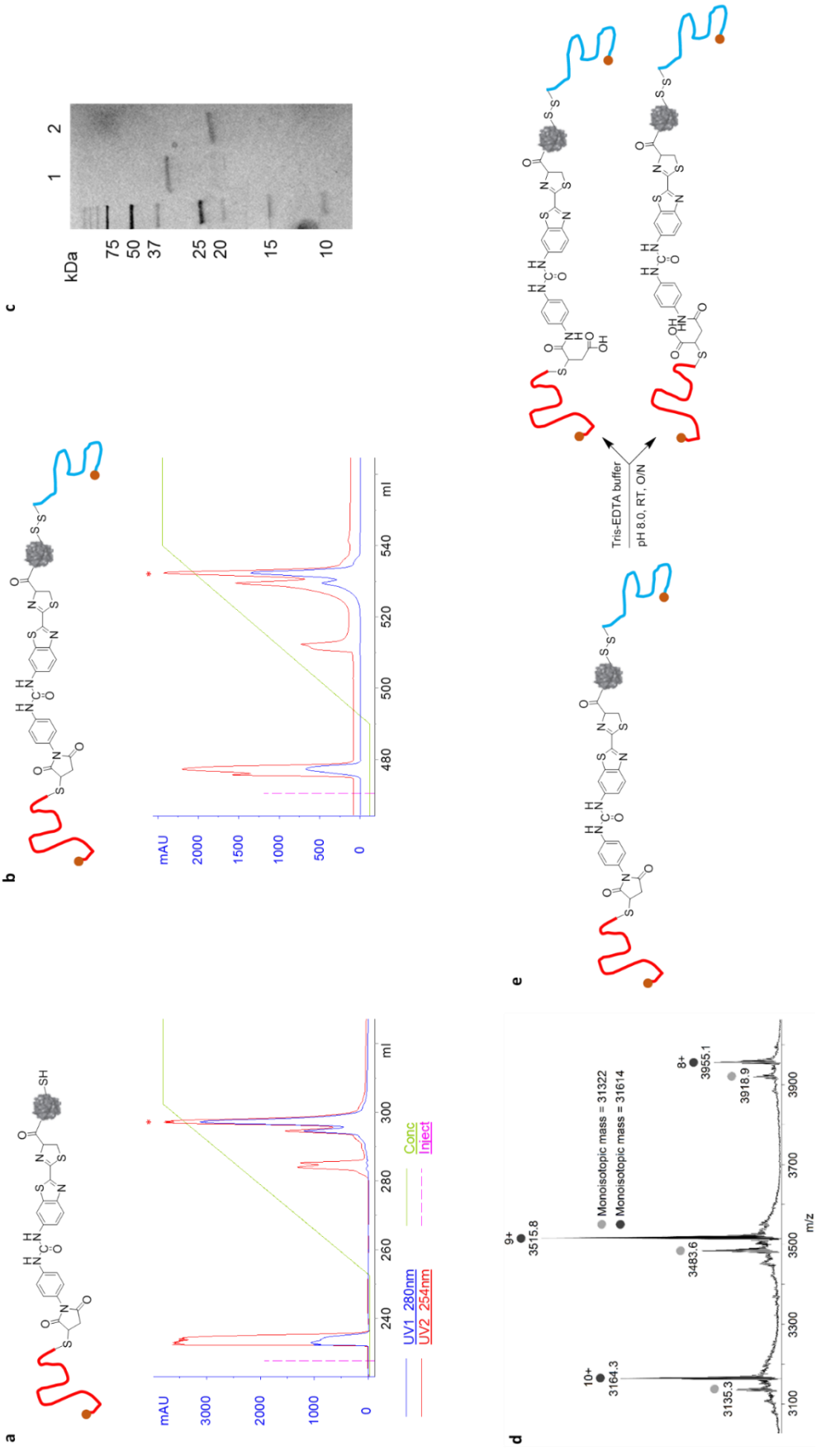


Figure 2.6 I Oligonucleotide-protein conjugates purification. Ion exchange chromatography purification of (a) btn-oligo$(dC)_{30}$-S1C-TrxV5-C109, (b) btn-oligo$(dC)_{30}$-S1C-TrxV5-C109-oligo$(dA)_{30}$-btn. The desired product is marked with a red asterisk. (c) SDS-PAGE gel: lane 1: btn-oligo$(dC)_{30}$-S1C-TrxV5-C109-oligo$(dA)_{30}$-btn, lane 2: btn-oligo$(dC)_{30}$-S1C-TrxV5-C109. (d) Native MS of btn-oligo$(dC)_{30}$-S1C-TrxV5-C109-oligo$(dA)_{30}$-btn. Calculated $[M \text{ (hydrolysis)} + H]^+ = 31598 + 18$, observed: 31614. Opening of the succinimide ring has occurred. The peaks corresponding to a mass of 31322 are derived from btn-oligo$(dC)_{30}$-S1C-TrxV5-C109-oligo$(dA)_{30}$-btn missing one C nucleotide (pdC). (e) In basic conditions, the ring formed by the maleimide can be hydrolyzed to give a mixture of regioisomers, but the S-C bond was not cleaved.

2.3 Discussion

In this chapter, we have developed a method for site-specific dual modification of protein (thioredoxin), with two different oligonucleotides. The N terminus conjugation was carried out with a bioorthogonal cyanobenzothiazole condensation, and the C terminus conjugation by disulphide bond formation. Advantages of this method include mild reaction conditions and high overall yield of the oligonucleotide-protein conjugates.

Notably, the cyanobenzothiazole reacts with the N-terminal cysteine site-specifically, and it works in the presence of internal or C-terminal cysteine. This efficient and mild N-terminal cysteine labelling method will allow finishing the protein target enriched from transfected cells after N-terminal cysteines have been revealed by enzymatic means.

The conjugate (btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn) developed here will be utilized for discriminating the direction of co-translocational protein unfolding in Chapter 3 and measuring the protein folding kinetics in Chapter 4.

CHAPTER 3

Discriminating the direction of co-translocational protein unfolding

3.1. Introduction

Protein transport between cell organelles, through transmembrane channels, is a ubiquitous biological process. The passage through these narrow pores is usually coincident with sequential unfolding, which initiates from one terminus [118,119]. For example, most mitochondrial proteins are synthesized in the cytosol. In mitochondria import, loosely folded precursor proteins with N- or C-terminal presequences [120] that contain mitochondrial-targeting signals are actively unfolded by the translocators to thread into their channels

(translocase of the outer membrane (TOM complex) and translocase of the inner membrane (TIM complex)) (Figure 3.1 a). The presequences then engage the import machineries (mitochondrial heat shock protein 70 (mtHsp70)) in the mitochondrial-matrix, and the mtHsp70s pull the proteins through the channels [121-124]. Chloroplast import occurs through a similar mechanism: unfolding of the terminal presequence-containing precursor protein upon transport into the chloroplast [125]. Moreover, in protein degradation, protein entry into the proteasome can be unidirectional or bidirectional, which is determined by the polyubiquitin tag and the stability of the termini (easily unfolded termini tend to translocate first) [110,126,127]. The proteasome then pulls the substrate, unfolding it as it enters the degradation chamber, where the protein is digested [99,110,128].

CHAPTER 3

Discriminating the direction of co-translocational protein unfolding

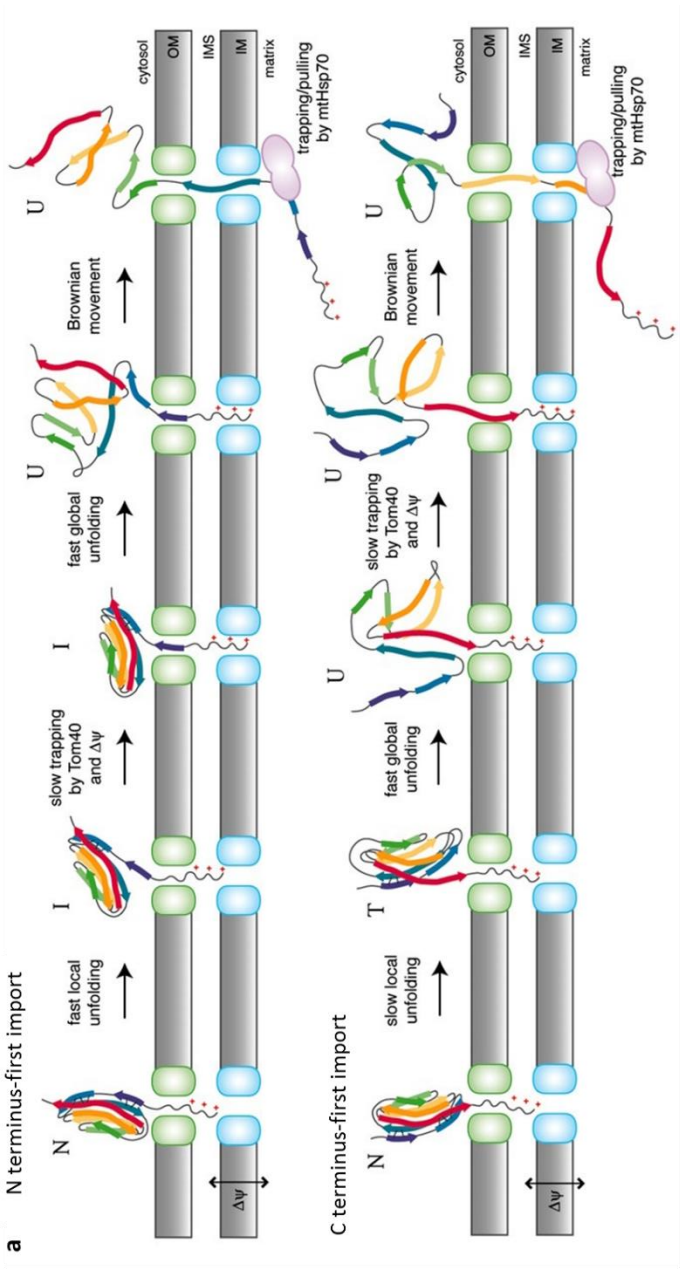


Figure 3.1 I Comparison of the protein unfolding pathways between mitochondria import and mechanical two-end pulling. (a) Protein unfolding and import into mitochondria from N terminus (upper panel) and C terminus (lower panel) led by a presequence. (b) Protein unfolding by using force spectroscopy. Protein: titin I27 domain (the 27th Ig domain of titin), N: native state, I: unfolding intermediate, T: transition state, U: unfolded state, OM: outer membrane, IM: inner membrane, IMS: intermembrane space, Δψ: membrane potential gradient. Adapted with permission from Sato *et al.* [109].

CHAPTER 3

Discriminating the direction of co-translocational protein unfolding

Experimentally, protein unfolding can be accelerated by global distortions with thermal or chemical denaturation, or mechanical two-end pulling using force spectroscopy. However, neither approach can elucidate the protein unfolding process during its directional channel transport (Figure 3.1) [109,129,130]. Stochastic nanopore sensors, based on a naturally occurring *Staphylococcal aureus* α-haemolysin (αHL) pore, have been developed with applications in nucleic acid analysis and DNA sequencing (Chapter 1, Section 1.2) [30-34]. The αHL pore is sufficiently versatile to provide information for other polymers such as polypeptides and proteins during their translocation (Chapter 1, Section 1.3), and therefore can be used to mimic and study this important biological phenomenon [83,84].

Recently, a small substrate, Smt3 domain, has been pulled, unfolded, and translocated through the αHL pore by using a molecular motor, the unfoldase ClpX (Chapter 1, Section 1.3.3) [96]. The translocation process was observed by analyzing the fluctuation of the ionic current through the pore when the peptide chain, detached from the unfolding protein, was translocated through the pore. This technique has also enabled the discrimination among protein variants [100]. Simultaneously, in another study, a model protein, thioredoxin, was unfolded during its forced translocation through an αHL pore by a covalently attached, negatively charged oligonucleotide leader sequence, which provides a tunable driving force of approximately 5–10 pN [42,104,105], under a positive applied potential (Chapter 1, Section 1.3.4) [97]. In this work, the pore-mediated thioredoxin unfolding has been revealed as four distinct steps with a partly unfolded intermediate, which is in a different unfolding

pathway for this paradigm of two-state unfolding observed in solution [107,108,116]. Further, when the protein was attached to the oligonucleotide at either N or C terminus and therefore unidirectionally unfolded from either terminus, the ionic current signatures produced during the translocation differed markedly (Chapter 1, Figure 1.9). The different translocation duration indicated different kinetics of co-translocational unfolding [106]. However, these initial studies that investigated the protein unidirectional translocation through the pore rely on the signatures established empirically for each terminus, and it is impossible to predict *a priori* the direction of translocation for a given protein in the case of bidirectional pulling.

The αHL pore has an inherent ability to detect nucleic acids (Chapter 1, Section 1.2.1) [33,35,36] and discriminate stationary oligonucleotides (Chapter 1, Section 1.2.3). It has been extensively observed that when the oligonucleotides were immobilized within the pore by a terminal hairpin [39,61] or a protein attached to the end [52-54], either of which are too large to allow the strand translocation, the stretched oligonucleotides within the pore produce a sequence- and orientation-dependent residual current. The prolonged observation time provided by immobilization improves the signal-to-noise ratio and thus allows the discrimination of small differences in the residual currents. Therefore, the translocation of protein pulled by the oligonucleotide can be advanced to analyze the direction of translocation.

To achieve such directionality analysis (N or C terminus-first), a thioredoxin variant bearing two different biotinylated oligonucleotide sequences at both N

and C termini was examined during its translocation through the αHL pore. The protein translocation led by one of the oligonucleotides, in this case, was bidirectional. Here, the oligonucleotides are not only responsible for the pulling of the protein unfolding and translocation [97], but are also expected to serve as terminal tags: the oligonucleotide tag attached to the terminus from which the protein is pulled should produce a distinguishable current blockade. The determined translocation direction, according to the ionic current signatures characterized previously [106], was corroborated by the residual currents produced by the two different oligonucleotide leader sequences (anchored by the thioredoxin prior to the unfolding) (Figure 3.5). Further, after the protein unfolding and translocation through the pore, the oligonucleotide follower sequence, attached to the other terminus of the protein, was immobilized by using streptavidin bound to its biotinylated end. This also provided the information confirming the direction of translocation (Figure 3.8). The direction of translocation analyzed by the “terminal oligonucleotide tag” strategy was consistent with previous reports for this model protein (thioredoxin) [106]. This transferable methodology is expected to be utilized to determine the direction of transport through the biological channel for any given protein analyte.

3.2 Results

3.2.1 Co-translocational protein unfolding pulled from N or C terminus

A thioredoxin variant, Trx V5 (C32S/C35S/A22P/I23V/P68A), was covalently attached to two biotinylated oligonucleotides, btn-oligo(dC)$_{30}$ and btn-oligo(dA)$_{30}$, at both N- and C-terminal cysteines (btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn) (Chapter 2). Under a positive voltage bias across the αHL pore, one of the anionic oligonucleotides, either N- or C-terminal, promotes the capture and retention of the protein terminus atop the pore entrance, and then pulls the protein unfolding and translocation through the pore (Figure 3.2 a and Figure 3.5 a, c).

The interactions between the oligonucleotide-protein conjugates and the pore generated a series of transient current blockades (Figure 3.2 b), which showed two similar types of repeating ionic current signatures (Figure 3.5 b, d). Those two patterns, with four current levels (an open-pore level and three blockade levels), represented N and C terminus-first protein unfolding and translocation through the pore [97,106]. Both signatures showed similar residual current values (I_{RES}) in all three current blockades [106]. The I_{RES} values reflect the extent to which the polymers (oligonucleotides or polypeptide chains) block the pore, and are involved in their conformation [52,53,96,97]. Therefore, as discussed before (Chapter 1, Section 1.3.4) [97,106] and for both N and C terminus-first translocations, level 2 represents a state where the thioredoxin is captured atop the *cis* entrance of the pore, and the oligonucleotide leader sequence is threaded spanning across the

pore. Level 3 represents a thioredoxin intermediate state, after partially unfolding from either terminus, with the detached polypeptide chain within the pore. Level 4 represents a state during which, after the protein is fully unfolded, the remainder of the polypeptide chain is held in the pore, without a potential driving force. A new cycle (level 1 to level 4) begins after the unfolded protein diffuses through the pore and the pore reassumes its open state (level 1).

The direction of thioredoxin translocation through the αHL pore is readily to be determined based on the ionic current signatures established in the case of unidirectional N and C terminus-first pulling in previous work (Chapter 1, Section 1.3.4) [106]. One of the most obvious features is that the current traces in level 3 showed strong differences in current noise, which is reflected in the width of their current histograms (Figure 3.3 a, b). In this experiment, the signals were low-pass-filtered at 5 kHz, and N terminus-first translocation produced a more noisy current trace compared with C terminus-first. This was also evident from the power spectral density analysis, where the spectra for N terminus-first showed higher values from the low frequencies up to the filter cutoff (5 kHz). Although the precise mechanism of the differences in current noise is unclear, this empirical observation is useful for determining the direction of translocation (for thioredoxin only). Moreover, for the current blockade duration, N terminus-first translocation was found 20–30 times shorter in level 2 (preunfolding state) but much longer in level 3 (intermediate state) than C terminus-first. (Figure 3.4). The different dwell-time distributions in each blockade level between N and C terminus-first indicated the different

co-translocational unfolding kinetics when pulling the protein from each terminus [106]. Aside from the characteristic ionic current signatures, we explored a new method of using a “terminal oligonucleotide tag” for discriminating the direction of protein translocation, which is discussed below.

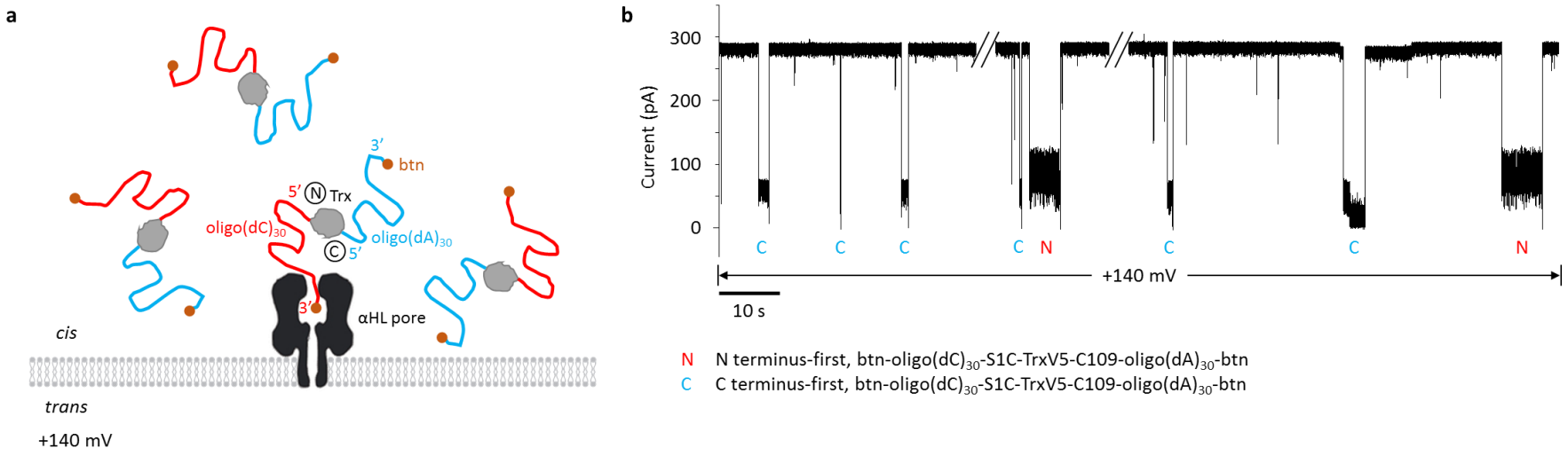


Figure 3.2 I Events of the conjugate (btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn) N and C terminus-first translocation through the αHL pore. (a) The bidirectional thioredoxin translocation through the pore can be facilitated from either N terminus pulled by oligo(dC)$_{30}$ or C terminus pulled by oligo(dA)$_{30}$. (b) Ionic current signals corresponding to N terminus-first (detail in Figure 3.5 b) or C terminus-first (detail in Figure 3.5 d) translocation. The applied voltage was +140 mV and the experiments were conducted at 22.9 °C.

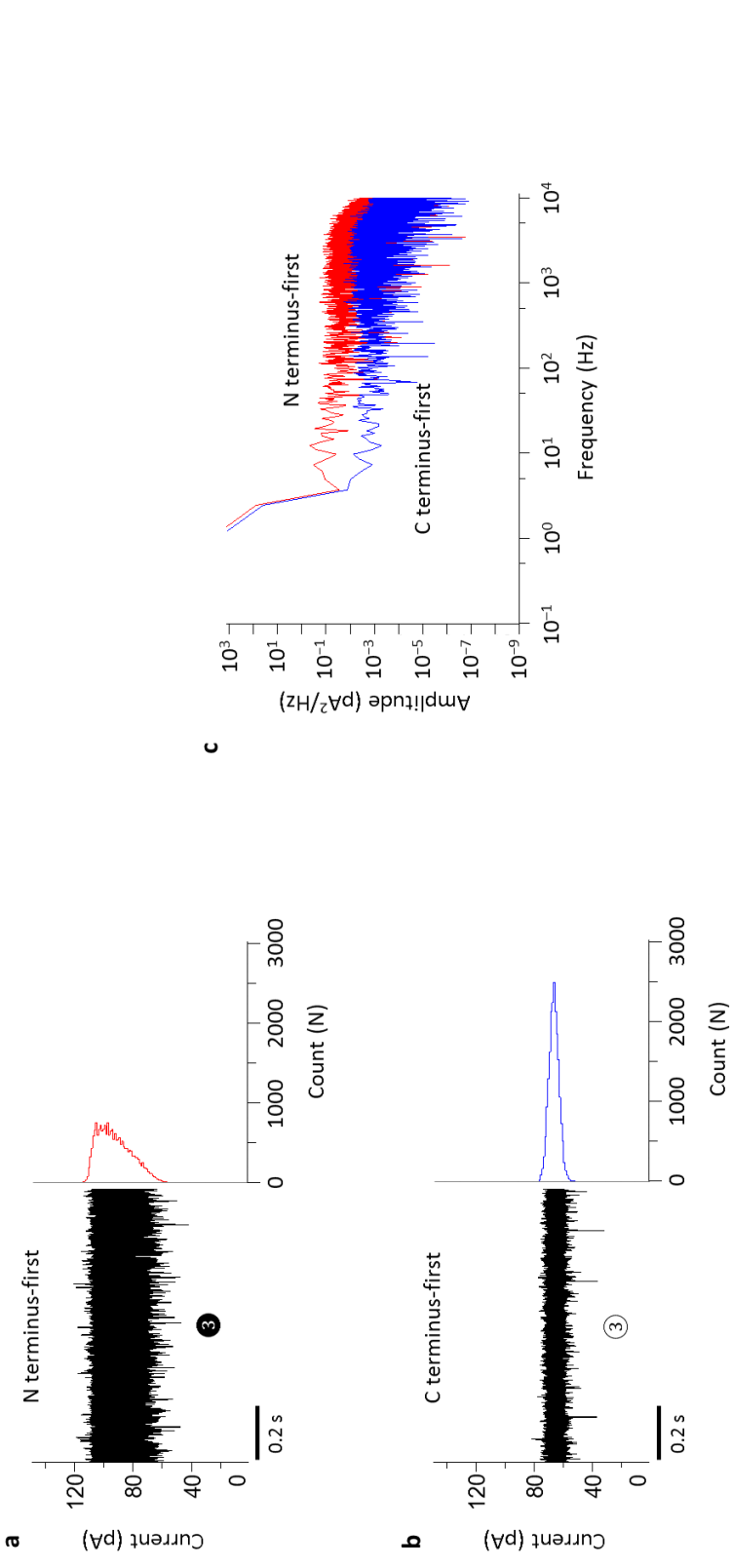


Figure 3.3 I Current noise of level 3 for N and C terminus-first translocation. Typical ionic current signals and histograms of level 3 for N terminus-first (a) and C terminus-first (b) translocation, illustrating clear differences in current noise. (c) Noise power spectral densities in level 3 for N and C terminus-first translocation. The applied voltage was +140 mV, and the current traces were filtered with a low-pass Bessel filter at 5 kHz. The experiments were conducted at 22.9 °C.

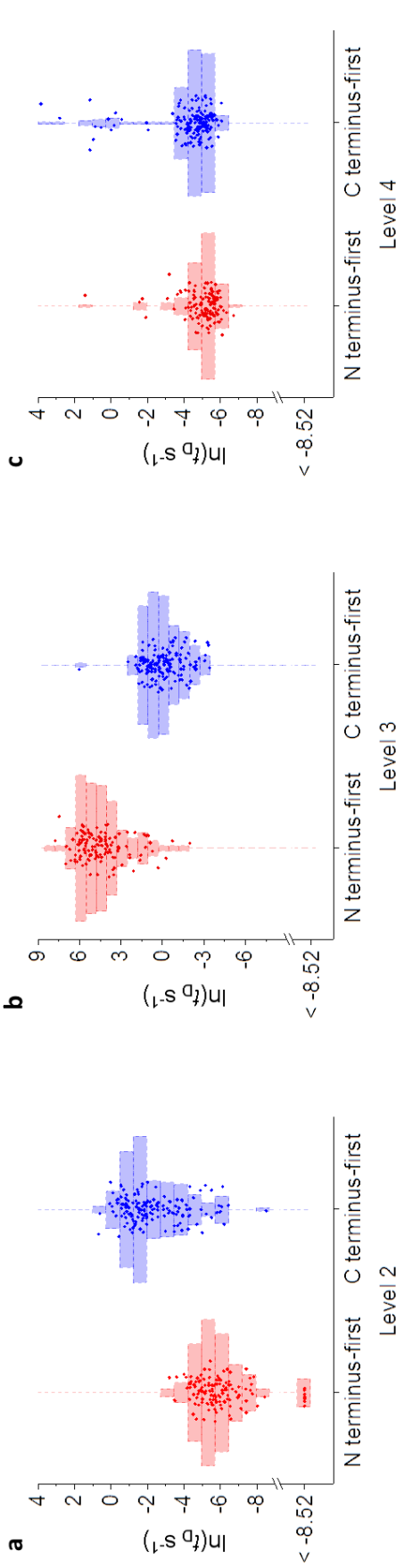


Figure 3.4 I Duration of each current level for N and C terminus-first translocation. Scatter plots of the dwell times in (a) level 2, (b) level 3 and (c) level 4 for N terminus-first (red) and C terminus-first (blue) translocation. Data were collected from three independent experiments. In total, 128 events for the N terminus-first and 160 events for the C terminus-first were analyzed. The applied voltage was +100 mV and the experiments were conducted at 22.9 °C.

3.2.2 Discriminating the direction of protein translocation by oligonucleotide leader sequence

The αHL pore can discriminate stationary oligonucleotides based on the sequence- and orientation-dependent current blockades [39,52-54,61]. At the state of level 2, the thioredoxin served as a steric "stopper" immobilizing the oligonucleotide (Figure 3.5 a, c). In this work, both termini of thioredoxin were covalently conjugated to two different oligonucleotides (btn-oligo$(dC)_{30}$-S1C-TrxV5-C109-oligo$(dA)_{30}$-btn): the N terminus was attached to the 5′ end of oligo$(dC)_{30}$ and the C terminus was attached to the 5′ end of oligo$(dA)_{30}$ (Chapter 2). Bidirectional translocation of thioredoxin through the pore can be facilitated from either the N terminus, led by oligo$(dC)_{30}$, or the C terminus, led by oligo$(dA)_{30}$, but in both cases, the oligonucleotide leader sequences are in the 3′ end-first orientation (Figure 3.2 a and Figure 3.5 a, c).

To test the feasibility of discriminating the direction of protein translocation by oligonucleotide leader sequence, we examined the I_{RES}% values in level 2 for N and C terminus-first ($I_{RES}^{N-, L2}$% and $I_{RES}^{C-, L2}$% respectively), and here we quote I_{RES}% as a percentage of the open-pore current (I_O). Three replicate experiments with different αHL pores were performed; the I_{RES}% values for N and C terminus-first translocation were readily distinguishable (Figure 3.6 and Table 3.1). Averaged over all three replicates, at +140 mV in 10 mM HEPES, 2 M KCl, pH 7.2, the αHL pores had a mean I_O of 291.52 ± 2.47 pA. The N terminus-first (oligo$(dC)_{30}$) gave: $I_{RES}^{N-, L2}$% = 16.52 ± 0.40%, while the C terminus-first (oligo$(dA)_{30}$) gave: $I_{RES}^{C-, L2}$% = 14.18 ± 0.18%. The residual

current difference between N and C-terminal translocation blockades in level 2 ($\Delta I_{RES}^{L2\%} = I_{RES}^{N-,\,L2}\% - I_{RES}^{C-,\,L2}\%$) was +2.35 ± 0.36%. These results agree well with those reported in the literature [52,53]: in the wild-type αHL pore and with the 3′ end entering the pore, the immobilized oligo$(dC)_{30}$ blocks the pore to a lesser extent than oligo$(dA)_{30}$ (summarized in Table 3.3). The distinguishable current blockade (level 2) suggested that the oligonucleotide leader sequence is sufficient to serve as a terminal tag to determine the direction of protein translocation through the αHL pore.

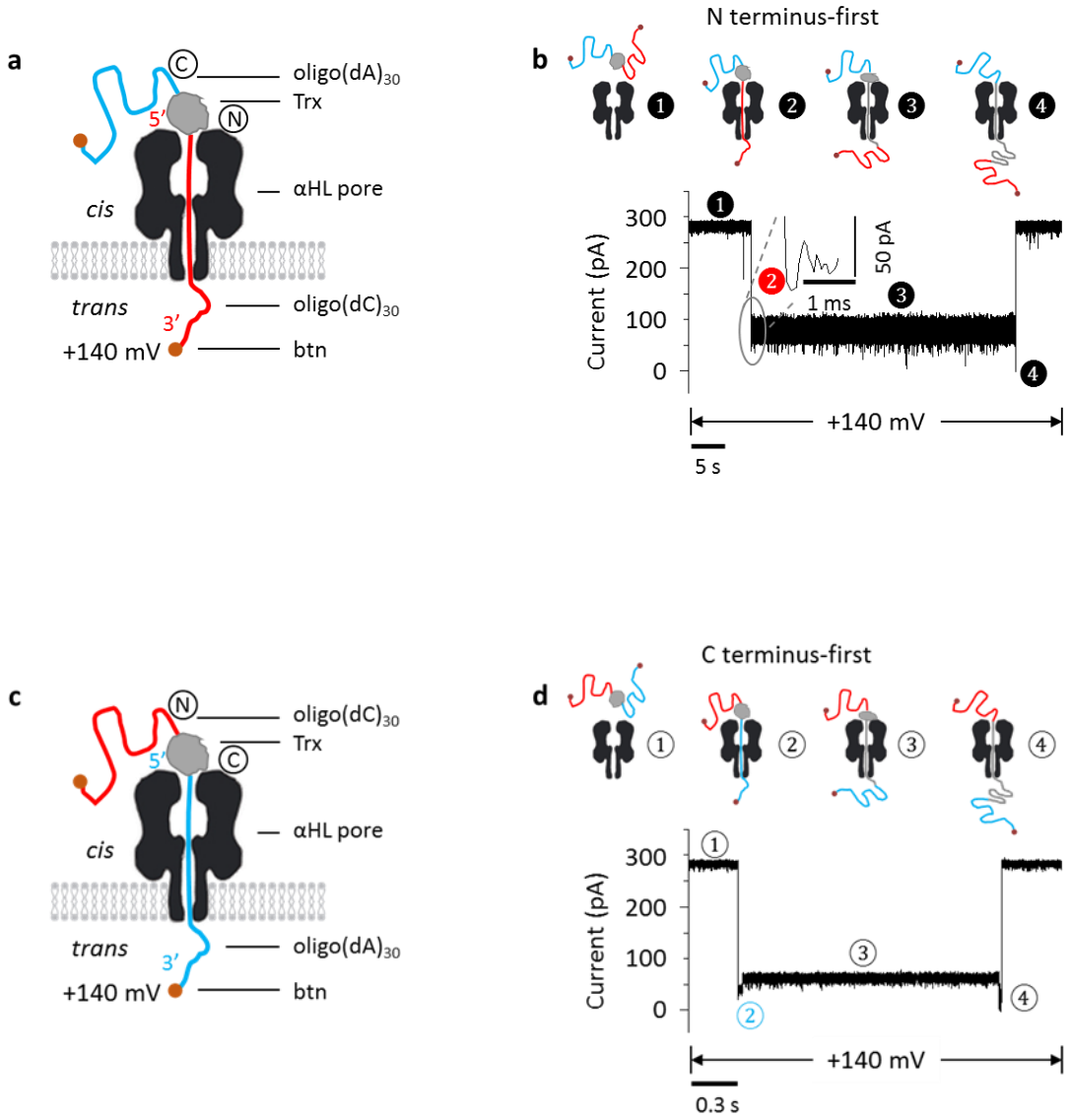


Figure 3.5 I Thioredoxin pulled bidirectional translocation through the αHL pore by oligonucleotide leader sequence. (a) Schematics depicting N terminus threading of the conjugate (btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn) into the *cis* entrance of the αHL pore, led by oligo(dC)$_{30}$. (b) Upper panel: N terminus-first translocation of the conjugate to the *trans* compartment of the pore. Lower panel: representative ionic current signal corresponding to N terminus-first translocation. (c) Schematics depicting C terminus threading of the conjugate into the *cis* entrance of the αHL pore, led by oligo(dA)$_{30}$. (d) Upper panel: C terminus-first translocation of the conjugate to the *trans* compartment of the pore. Lower panel: representative ionic current signal corresponding to C terminus-first translocation.

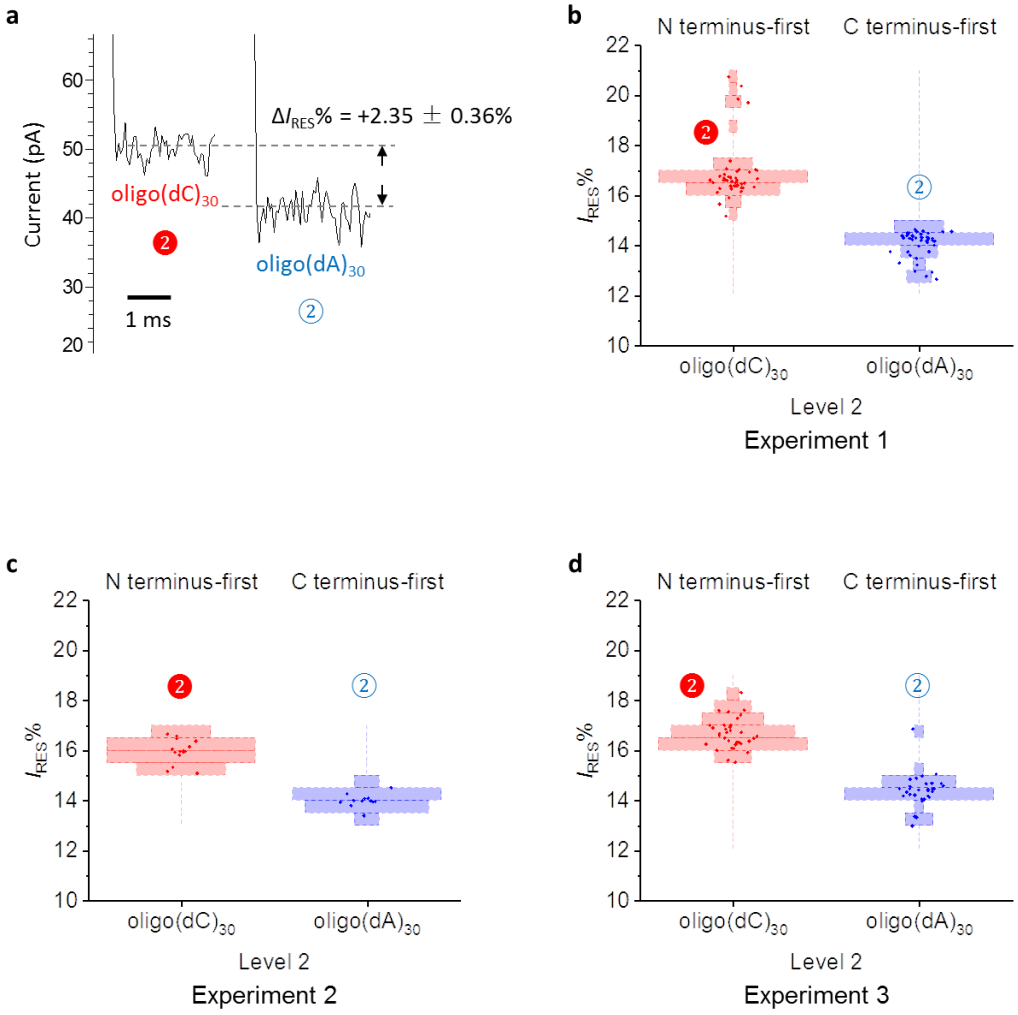


Figure 3.6 I Residual current of level 2 for N and C terminus-first translocation. (a) Typical ionic current signals of level 2 for N and C terminus-first translocation, respectively, corresponding to states with oligonucleotide leader sequences, oligo(dC)$_{30}$ and oligo(dA)$_{30}$, located within the αHL pore. (b-d) Scatter plots of I_{RES}% values of level 2 for N terminus-first (oligo(dC)$_{30}$, red) and C terminus-first (oligo(dA)$_{30}$, blue) translocation from three independent experiments. In total, 93 events for the N terminus-first and 87 events for the C terminus-first were analyzed; ~ 10% events for the N terminus-first were ignored as the current level 2 was too short (< 0.2 ms) for analysis. The applied voltage was +140 mV and the experiments were conducted at 22.9 °C.

Table 3.1 l I_{RES}% values of level 2 and level 5 for N and C terminus-first translocation

Experiments	I_O, pA	$I_{RES}^{N-, L2}$%	$I_{RES}^{C-, L2}$%	ΔI_{RES}^{L2}%	$I_{RES}^{N-, L5}$%	$I_{RES}^{C-, L5}$%	ΔI_{RES}^{L5}%
1	293.61	16.92	14.10	+2.83	17.80	17.45	+0.35
2	288.80	15.98	14.01	+1.97	17.28	16.91	+0.37
3	292.15	16.67	14.43	+2.24	17.83	17.50	+0.33
Mean (S.D.)	291.52 ± 2.01	16.52 ± 0.40	14.18 ± 0.18	+2.35 ± 0.36	17.64 ± 0.25	17.29 ± 0.27	+0.35 ± 0.02

The I_O and I_{RES}% values given for each level are mean values taken from Gaussian fits to event histograms for individual experiments. ΔI_{RES}% is defined as the difference in residual current between the N and C terminus-first translocations in level 2 ($\Delta I_{RES}^{L2}\% = I_{RES}^{N-, L2}\% - I_{RES}^{C-, L2}\%$) or level 5 ($\Delta I_{RES}^{L5}\% = I_{RES}^{N-, L5}\% - I_{RES}^{C-, L5}\%$).

3.2.3 Discriminating the direction of protein translocation by oligonucleotide follower sequence

As with the oligonucleotide leader sequence corroborating the direction of translocation when the protein enters the pore (Section 3.2.2), the oligonucleotide sequence attached to the other terminus (follower sequence) was also expected to produce distinguishable current blockade when the protein leaves through the pore. However, after the unfolded protein diffuses through the pore in step 4→1 (transition between level 4 and level 1), the oligonucleotide follower sequence freely translocates, and this process was too fast to be observed [30,79]. To detect the oligonucleotide follower sequence and further corroborate the direction of protein translocation, we used the biotin-streptavidin immobilization strategy, which was to bind the biotin tag at the end of one of the oligonucleotides to a high-affinity protein tag, streptavidin [52-54]. Streptavidin is unable to fit in the αHL pore and can be used to anchor the oligonucleotide within the pore. In this work, we used monovalent streptavidin, which is a streptavidin variant with a single femtomolar biotin binding site [65]. Each molecule can bind a single biotinylated oligonucleotide, so as to avoid the interference from the others.

In the experiment, the conjugate (btn-oligo$(dC)_{30}$-S1C-TrxV5-C109-oligo$(dA)_{30}$-btn) with biotin tags at the 3′ end of the oligonucleotide was allowed to form complexes with the monovalent streptavidin. We incubated an approximately two-fold stoichiometric excess of the conjugate with the monovalent streptavidin (mSA), which yielded mixtures of dual-bound

constructs (mSA-btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn-mSA), mono-bound constructs (btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn-mSA or mSA-btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn) and non-bound constructs (btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn) (Figure 3.7 a). When performing the protein translocation analysis, the dual-bound constructs were not able to enter the pore; non-bound constructs could translocate bidirectionally from either N or C terminus as described in section 3.2.2. The mono-bound constructs could translocate unidirectionally from either N terminus (btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn-mSA) (Figure 3.8 a) or C terminus (mSA-btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn) (Figure 3.8 c) respectively, and produce a five-level ionic current signature at a positive voltage (Figure 3.7 b and Figure 3.8 b, d). As demonstrated in the preceding text (Section 3.2.1), current level 2 to level 4 represent the protein unfolding and translocation through the pore, and the direction of translocation can be determined based on these well-characterized current levels. The newly observed stable current level 5 would be produced by the oligonucleotide follower sequence immobilized by the monovalent streptavidin at the *cis* entrance of the pore, and it can be held in that configuration as long as the applied voltage is maintained. To restore the pore to its open state, the voltage was reversed, and the thioredoxin was pulled retro-translocating back through the pore (current level 6) and left through the pore from the entrance, as the oligonucleotide leader sequence had no monovalent streptavidin attached (Figure 3.8 b, d).

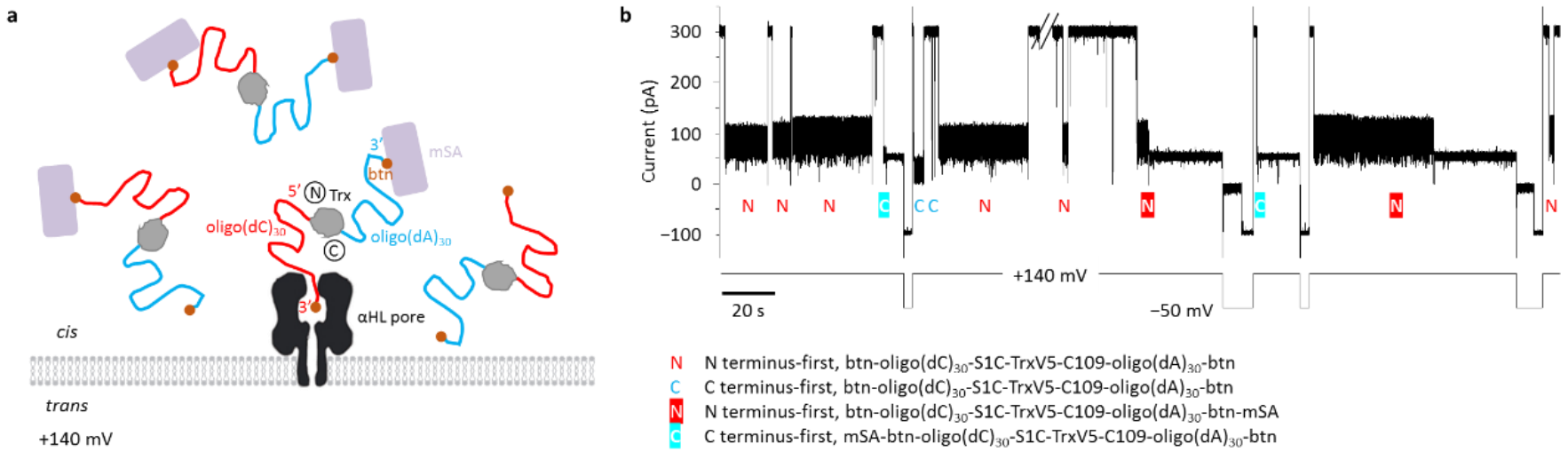


Figure 3.7 I Events of the constructs N and C terminus-first translocation and retro-translocation through the αHL pore. (a) The dual-bound construct (mSA-btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn-mSA) is not able to enter the pore; non-bound construct (btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn) can translocate bidirectionally from either N or C terminus (as described in Figure 3.2). The mono-bound construct can translocate unidirectionally from either N terminus (btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn-mSA) or C terminus (mSA-btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn). (b) ionic current signals corresponding to N terminus-first (detail in Figure 3.5 b and Figure 3.8 b) or C terminus-first (detail in Figure 3.5 d and Figure 3.8 d) translocation and retro-translocation. The applied voltage was +140 mV (translocation) and −50 (retro-translocation), and the experiments were conducted at 22.9 °C.

According to the designed construct, the statically captured oligonucleotide follower sequences could be oligo(dA)$_{30}$ in the case of N terminus-first translocation (Figure 3.8 a) or oligo(dC)$_{30}$ in the C terminus-first (Figure 3.8 c); in both cases, they were in the 5′ end-first orientation. We compared the I_{RES}% values in level 5 between N and C terminus-first ($I_{RES}^{N-, L5}$% and $I_{RES}^{C-, L5}$% respectively) in three different pores (Figure 3.9). The N terminus-first (oligo(dA)$_{30}$) gave: $I_{RES}^{N-, L5}$% = 17.64 ± 0.25%, while the C-terminus-first (oligo(dC)$_{30}$) gave: $I_{RES}^{C-, L5}$% = 17.29 ± 0.27%. The residual current difference between N- and C-terminal translocation blockades in level 5 (ΔI_{RES}^{L5}% = $I_{RES}^{N-, L5}$% – $I_{RES}^{C-, L5}$%) was +0.35 ± 0.02% (averaged over three replicates) (Table 3.1). It should be noted that the I_{RES}^{L5}% values showed experimental variation, nevertheless, in each individual experiment, the small ΔI_{RES}^{L5}% values were readily distinguishable. These results are consistent with previous work that, in the streptavidin-anchored experiment, oligo(dC)$_{30}$ blockades give a lower I_{RES}% than oligo(dA)$_{30}$ in the 5′ end-first orientation (summarized in Table 3.3) [52,53].

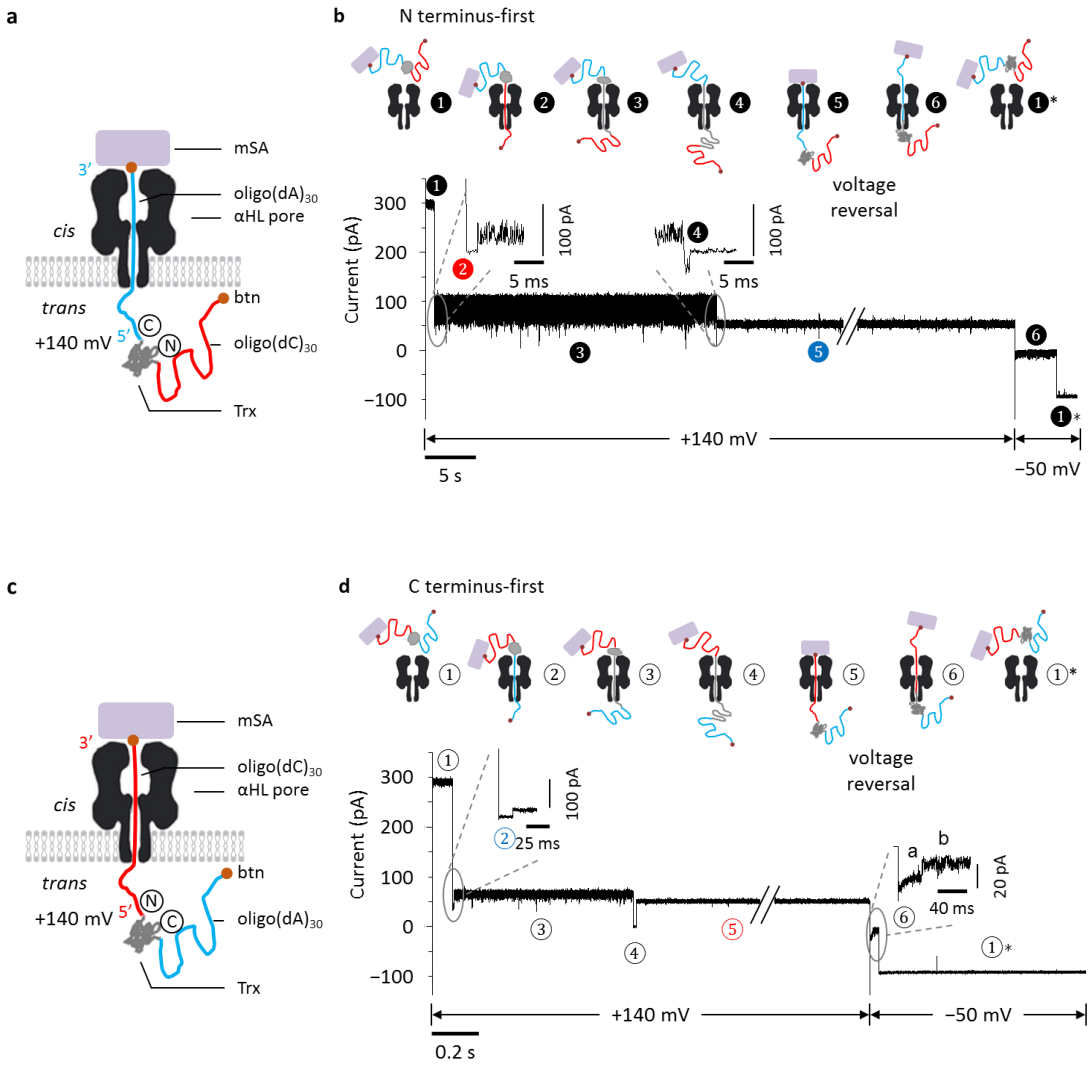


Figure 3.8 I Thioredoxin pulled bidirectionally through the αHL pore with the oligonucleotide follower sequence immobilized within the pore. (a) Schematics depicting N terminus-first translocation of the construct (btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn-mSA) through the αHL pore with oligo(dA)$_{30}$ immobilized within the pore. The C terminus is bound to a monovalent streptavidin. (b) Upper panel: N terminus-first translocation of the construct to the *trans* compartment of the pore, followed by retro-translocation to the *cis* compartment. Lower panel: representative ionic current signal corresponding to N terminus-first translocation and retro-translocation. (c) Schematics depicting C terminus-first translocation of the construct (mSA-btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn) through the αHL pore with oligo(dC)$_{30}$ immobilized within the pore. The N terminus is bound to a monovalent streptavidin. (d) Upper panel: C terminus-first translocation of the construct to the *trans* compartment of the pore, followed by retro-translocation to the *cis* compartment. Lower panel: representative ionic current signal corresponding to C terminus-first translocation and retro-translocation.

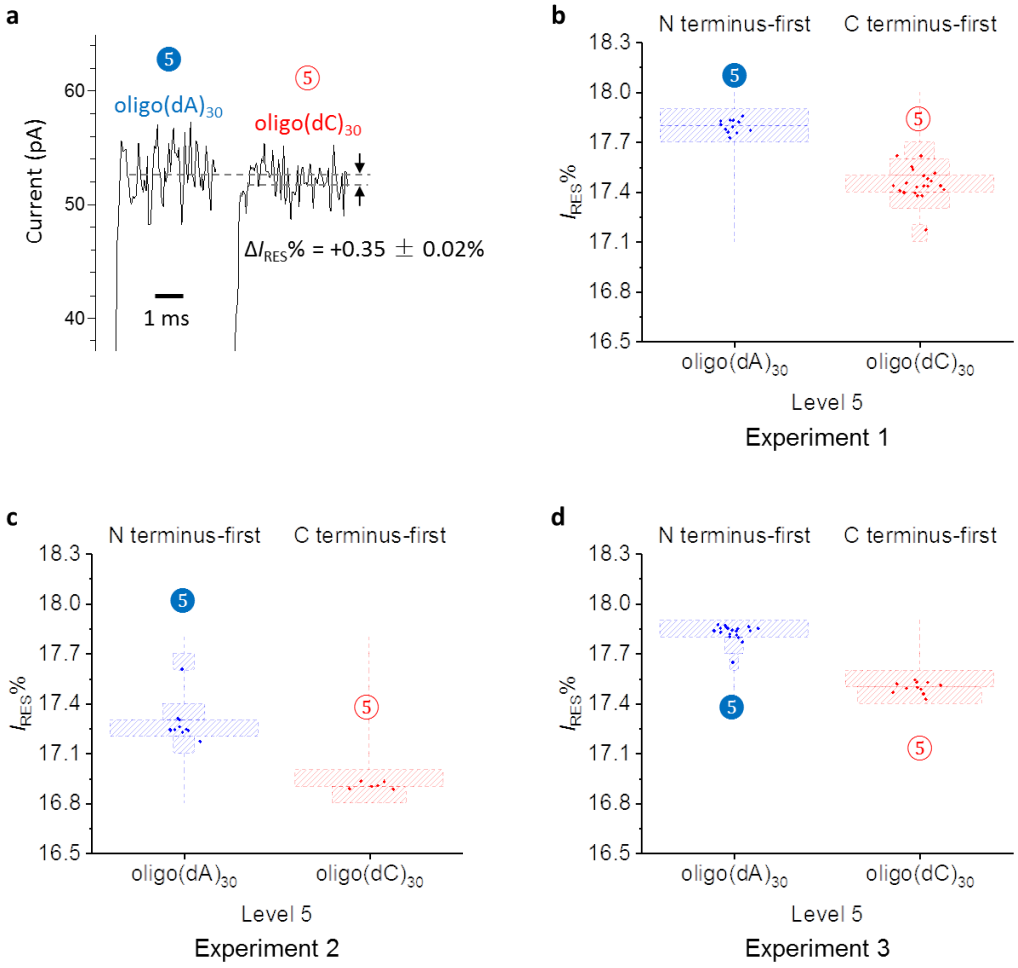


Figure 3.9 I Residual current of level 5 for N and C terminus-first translocation. (a) Typical ionic current signals of level 5 for N and C terminus-first translocation, respectively, corresponding to states with oligonucleotide follower sequences, oligo(dA)$_{30}$ and oligo(dC)$_{30}$, located within the αHL pore. (b-d) Scatter plots of I_{RES}% values of level 5 for N terminus-first (oligo(dA)$_{30}$, blue) and C terminus-first (oligo(dC)$_{30}$, red) translocation from three independent experiments. In total, 42 events for the N terminus-first and 38 events for the C terminus-first were analyzed. the applied voltage was +140 mV and the experiments were conducted at 22.9 °C.

In addition to the residual current (I_{RES}%), current noise (I_n) analysis can also be achieved due to the prolonged observation time in level 5. I_n is the standard deviation of a Gaussian fit to all-points histogram for the current

blockade and illustrates the current noise level (Figure 3.10 a, b). To accurately analyse the current noise, the current traces were recorded for 10 s. The N terminus-first translocation (oligo(dA)$_{30}$) gave higher I_n values ($I_n^{N-, L5}$ = 2.41 ± 0.18 pA) than the C terminus-first (oligo(dC)$_{30}$) ($I_n^{C-, L5}$ = 1.68 ± 0.16 pA) (averaged over three replicates) (Figure 3.10 d-f and Table 3.2). Power spectral density analysis revealed that the N terminus-first exhibited higher values from the low frequencies up to the filter cutoff (5 kHz) than the C terminus-first (Figure 3.10 c). These values indicated that the statically captured oligo(dA)$_{30}$ produced a noisy current trace than oligo(dC)$_{30}$ in the 5′ end-first orientation.

The distinct differences in current level 5, produced by monovalent streptavidin-anchored oligo(dA)$_{30}$ and oligo(dC)$_{30}$ within the αHL pore, suggested that the oligonucleotide follower sequence, can also be used as a supplement to discriminate the direction of protein translocation.

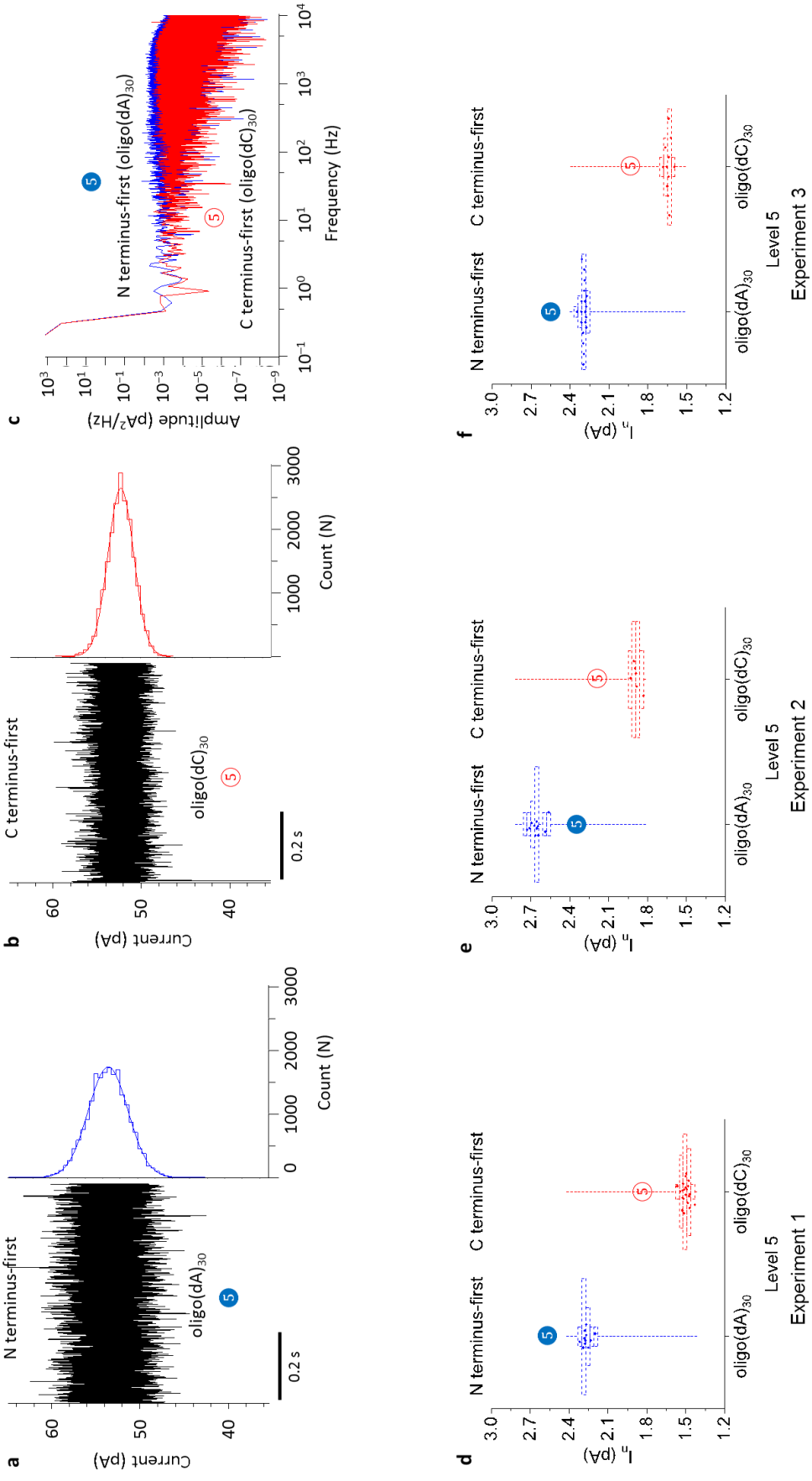

a
N terminus-first
oligo(dA)30
5
0.2 s
Current (pA)
Count (N)
b
C terminus-first
oligo(dC)30
5
0.2 s
Current (pA)
Count (N)
c
N terminus-first (oligo(dA)30)
C terminus-first (oligo(dC)30)
5
5
Amplitude (pA²/Hz)
Frequency (Hz)
d
N terminus-first
C terminus-first
5
5
oligo(dA)30
oligo(dC)30
In (pA)
Level 5
Experiment 1
e
N terminus-first
C terminus-first
5
5
oligo(dA)30
oligo(dC)30
In (pA)
Level 5
Experiment 2
f
N terminus-first
C terminus-first
5
5
oligo(dA)30
oligo(dC)30
In (pA)
Level 5
Experiment 3

Figure 3.10 l Current noise of level 5 for N and C terminus-first translocation. Typical ionic current signals and histograms of level 5 for N terminus-first (a) and C terminus-first (b) translocation, respectively, corresponding to states with oligonucleotide follower sequences, oligo(dA)$_{30}$ and oligo(dC)$_{30}$, located within the αHL pore. (c) Noise power spectral densities in level 5 for N terminus-first (oligo(dA)$_{30}$, blue) and C terminus-first (oligo(dC)$_{30}$, red) translocation. (d-f) Scatter plots of I_n values of level 5 for N terminus-first (oligo(dA)$_{30}$, blue) and C terminus-first (oligo(dC)$_{30}$, red) translocation from three independent experiments. In total, 42 events for the N terminus-first and 38 events for the C terminus-first were analyzed. The applied voltage was +140 mV, and the current traces were filtered with a low-pass Bessel filter at 5 kHz. The experiments were conducted at 22.9 °C.

Table 3.2 l I_n values of level 5 for N and C terminus-first translocation

Experiments	$I_n^{N-, L5}$, pA	$I_n^{C-, L5}$, pA	ΔI_n^{L5}, pA
1	2.27	1.50	0.77
2	2.66	1.89	0.77
3	2.29	1.64	0.65
Mean (S.D.)	2.41 ± 0.18	1.68 ± 0.16	+0.73 ± 0.05

The I_n values given for level 5 are mean values taken from Gaussian fits to event histograms for individual experiments. ΔI_n^{L5} is defined as the difference in noise between the N and C terminus-first translocations in level 5 ($\Delta I_n^{L5} = I_n^{N-, L5} - I_n^{C-, L5}$).

3.3 Discussion

Considerable efforts using ensemble techniques have been devoted to address the fundamental biological process of protein unfolding during its directional transport through transmembrane channels [99,109,110,121,128,129]. Recent nanopore experiments using a potential-driven strategy have shed light on the fine details of a model protein (thioredoxin) unidirectionally, co-translocationally unfolding (N or C terminus-first) and diffusing through a model pore (αHL pore), and showed that the protein unfolding kinetics differ depending on which terminus of the substrate is threaded [97,106]. Here, we have taken advantage of the ability of the αHL pore to discriminate stationary oligonucleotides to advance the potential-driven method for protein translocation, which enables the discrimination of the translocation direction in the case of bidirectional pulling.

In summary, we demonstrated that, in the process of site-specific dual-tagged thioredoxin (btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn) bidirectional translocation through the αHL pore, different current levels between N and C terminus-first were produced by the oligonucleotide tags when the protein enters (level 2: thioredoxin-anchored oligonucleotide leader sequences) and leaves through (level 5: monovalent streptavidin-anchored oligonucleotide follower sequences) the pore. Therefore, analysis of the residual current values of those “terminal oligonucleotide tags” reveals the direction of protein co-translocational unfolding, with no requirement of prior knowledge of the distinctive ionic current signature produced by the protein. This methodology

is thus expected to be applied to any protein analyte for the directional nanopore translocation analysis.

The residual current levels for the αHL pore containing homo-oligonucleotides, in various conditions, have been systematically investigated before [37,52-54,61]. The residual current values obtained in this work are in approximate agreement with these previous works (summarized in Table 3.3). The small differences between batches could be attributed to the different external conditions. Experimental settings (engineered pore, oligonucleotide immobilization technique) and external conditions (buffer solution, pH, salt concentration, temperature and applied voltage) can affect the values. It is worth noting that an αHL pore variant (E111N/K147N) achieves a significantly larger residue current difference between the streptavidin-anchored oligonucleotides (oligo(dA) and oligo(dC)) in the 5' end-first orientation [53]. It will be interesting to use this engineered pore to characterize the ionic current signature during the co-translocational unfolding of thioredoxin and discriminate the direction of translocation: better discrimination between oligonucleotide follower sequences (level 5) can be attained, but different current signals will be produced because of the different pore interior geometry (Figure 3.11 and Figure 3.12).

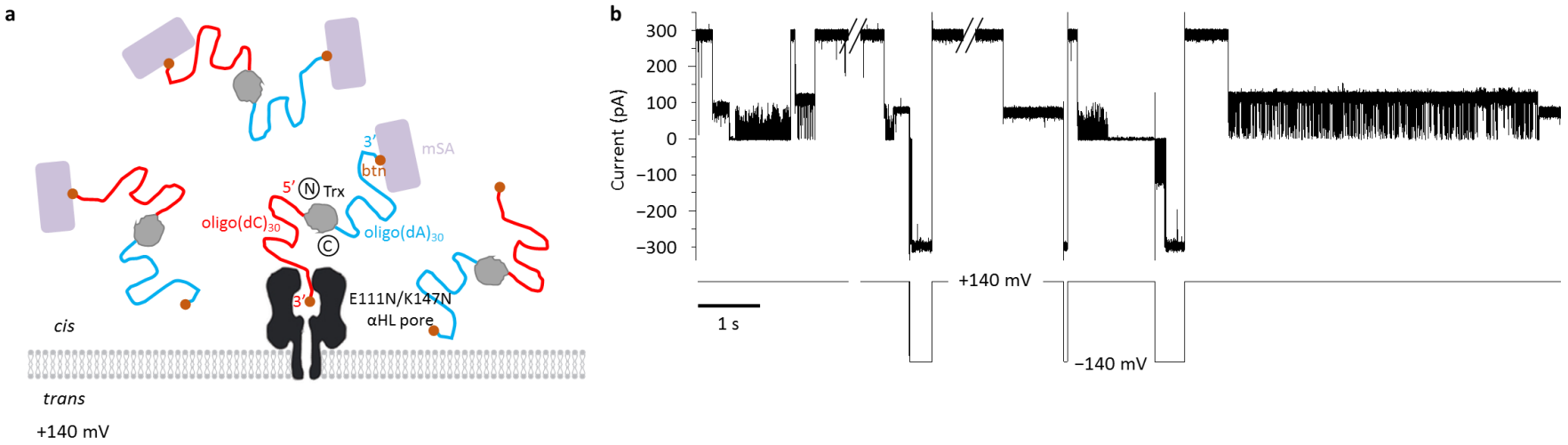


Figure 3.11 I Events of the constructs translocation and retro-translocation through the E111N/K147N αHL pore. (a) The dual-bound construct (mSA-btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn-mSA) is not able to enter the pore; the non-bound construct (btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn) can translocate bidirectionally. The mono-bound constructs (btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn-mSA and mSA-btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn) can translocate unidirectionally. (b) Ionic current signals corresponding to the constructs translocation (and retro-translocation) (detail in Figure 3.12). Events of the constructs translocation (and retro-translocation) through the wild-type αHL pore were described in Figure 3.2 and Figure 3.7. The applied voltage was +140 mV (translocation) and −140 (retro-translocation), and the experiments were conducted at 22.9 °C.

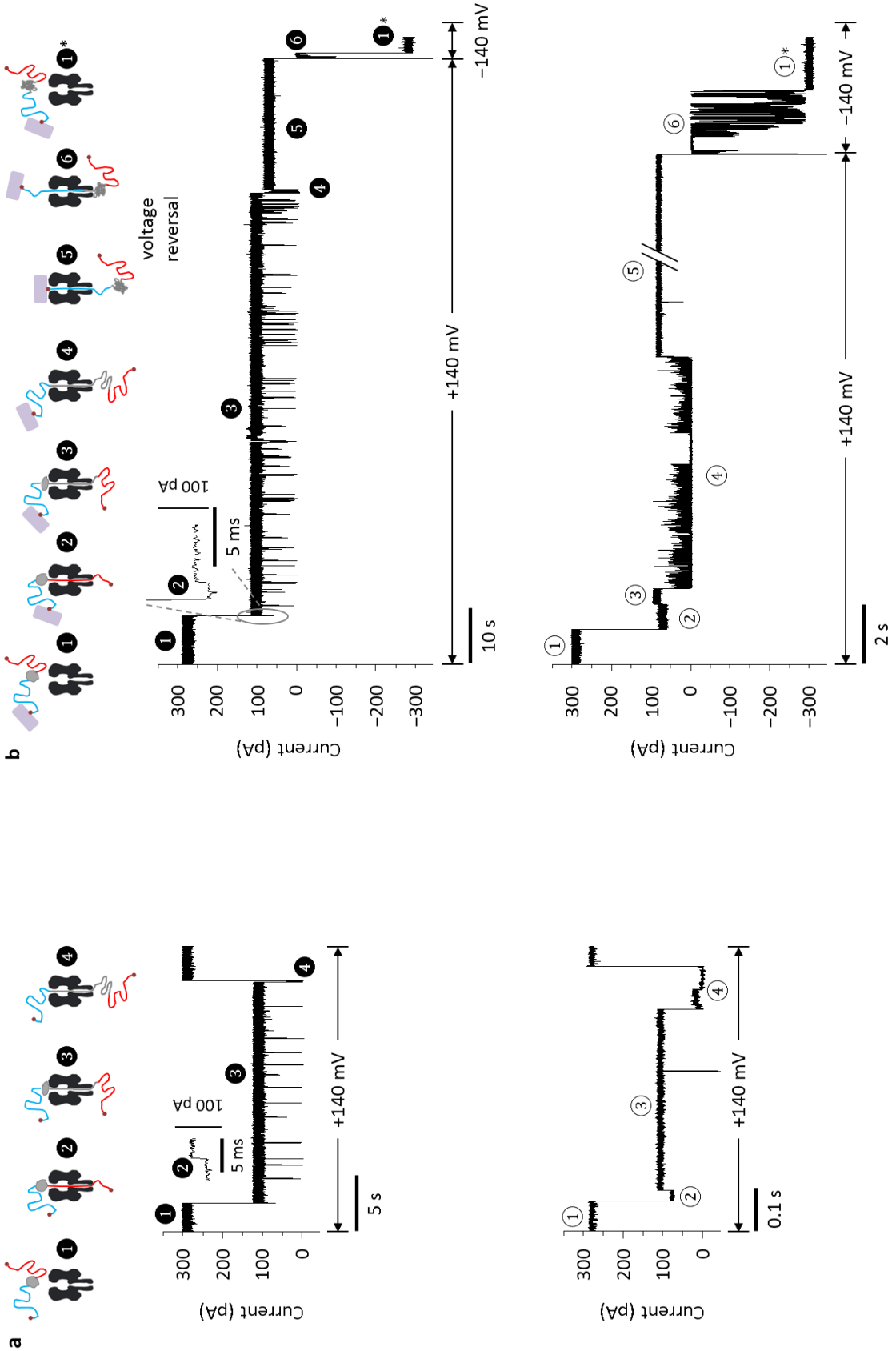

a
b
voltage reversal
+140 mV
−140 mV
5 ms
100 pA
10 s
5 s
2 s
0.1 s
Current (pA)
300
200
100
0
−100
−200
−300

Figure 3.12 I Four types of ionic current signals of the constructs translocation (and retro-translocation) through the E111N/K147N αHL pore. (a) Upper panel: translocation of the non-bound construct (btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn) to the *trans* compartment of the E111N/K147N αHL pore. Lower panel: two representative types of ionic current signal of the non-bound construct translocation (N and C terminus-first translocation). (b) Upper panel: translocation of the mono-bound construct (btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn-mSA or mSA-btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn) to the *trans* compartment of the E111N/K147N αHL pore, followed by retro-translocation to the *cis* compartment. Lower panel: two representative types of ionic current signal of the mono-bound construct translocation and retro-translocation (N and C terminus-first translocation). The proposed current levels that reflect the constructs translocation through the E111N/K147N αHL pore (level 1 to level 4), oligonucleotide follower sequence immobilized within the pore (level 5), and retro-translocation through the pore (level 6 to level 1*) were indicated in each ionic current signal. The characteristic ionic current signals of these events with the wild-type αHL pore were described in Figure 3.5 b, d and Figure 3.8 b, d.

Table 3.3 l I_{RES}% values for αHL pore containing oligo(dA) and oligo(dC)

Oligonucleotide immobilization technique	αHL pore	Threaded orientation	$I_{RES}^{oligo(dA)}$%	$I_{RES}^{oligo(dC)}$%	ΔI_{RES}%	Conditions (buffer / applied voltage)	Reference
thioredoxin	wild-type	3’	14.2	16.5	−2.3	2 M KCl, 10 mM HEPES, pH 7.2 / +140 mV	current work
monovalent streptavidin	wild-type	5’	17.6	17.3	+0.3	2 M KCl, 10 mM HEPES, pH 7.2 / +140 mV	current work
streptavidin	wild-type	3’	16.0	18.9	−2.9	1 M KCl, 10 mM Tris·HCl, 1 mM EDTA, pH 8.0 / +120 mV	[52]
streptavidin	wild-type	5’	18.3	17.1	+1.2	1 M KCl, 10 mM Tris·HCl, 1 mM EDTA, pH 8.0 / +120 mV	[52]
streptavidin	wild-type	5’	20.0	19.4	+0.6	1 M KCl, 25 mM Tris·HCl, 0.1 mM EDTA, pH 8.0 / +160 mV	[53]
streptavidin	E111N/K147N	5’	33.9	36.6	−2.7	1 M KCl, 25 mM Tris·HCl, 0.1 mM EDTA, pH 8.0 / +160 mV	[53]
terminal hairpin	wild-type	3’	22.0	31.0	−9.0	500 mM KCl, 5 mM MOPS, pH 7.5 / +170 mV	[61]
terminal hairpin	wild-type	5’	21.5	32.0	−10.5	500 mM KCl, 5 mM MOPS, pH 7.5 / +170 mV	[61]
N/A (freely translocating)	wild-type	unknown	12.6	13.4	−0.8	1 M KCl, 10 mM Tris·HCl, pH 8.5 / +120 mV	[37]

ΔI_{RES}% is defined as the difference in residual current between the oligo(dA) and oligo(dC) blockades (ΔI_{RES}% = $I_{RES}^{oligo(dA)}$% – $I_{RES}^{oligo(dC)}$%). The experiments were conducted between 19 °C and 25 °C.

CHAPTER 4

Single-molecule measurement of protein folding kinetics with a nanopore

4.1. Introduction

While protein unfolding is commonplace in cellular activity, its reversal, folding, is also a fundamental process. A polypeptide's amino acid sequence and interactions with itself determine its native folded structure [131]. In the crowded cellular environment, the complex solvent (solutes) largely affect this process [132-134]. Folding happens first upon protein ribosomal synthesis: it initiates co-translationally from the nascent polypeptide N terminus as it is synthesized and emerges from the ribosomal exit tunnel (Figure 4.1) [135-

137]. The spatial constrains within the ribosome tunnel influence the co-translational folding route [135,136,138], and auxiliary proteins, such as ribosome-associated chaperones, interact with incompletely folded chains so as to favor successful folding and disfavor aggregation [139]. In addition, proteins destined for secretion (secretory protein) are synthesized on the endoplasmic reticulum-bound ribosomes; folding occurs during the co-translational translocation through a membrane-conducting channel into the endoplasmic reticulum [140,141].

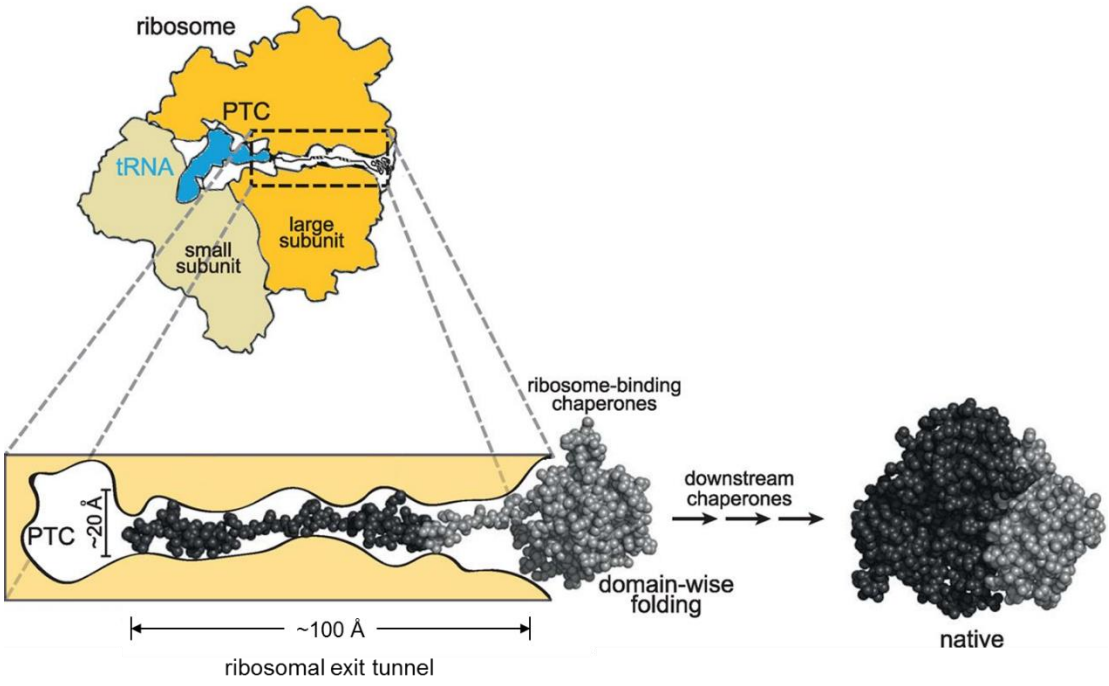


Figure 4.1. Protein folding at the ribosomal exit tunnel. The exit tunnel of the large ribosomal subunit, as seen in a longitudinal cut through the ribosome, is magnified and shown schematically. The nascent polypeptide chain of a multidomain protein is modeled into the tunnel, forming a folded N-terminal domain immediately outside the ribosomal exit tunnel. This co-translational folding process may be assisted by ribosome-associated chaperones. PTC, peptidyl transferase center. Adapted with permission from Balchin *et al*. [134].

CHAPTER 4

Single-molecule measurement of protein folding kinetics with a nanopore

In the course of protein trafficking between cell organelles, such as protein import into mitochondria, folded proteins are unfolded and transported through nanoscale channels (translocase of the outer membrane (TOM complex) and translocase of the inner membrane (TIM complex)) (Chapter 3, Figure 3.1) [121-124]. The proteins then refold into biologically active conformations in the recipient compartment (mitochondrial-matrix). Refolding of the translocated protein starts from the newly imported polypeptide chain protruding from the end of the channel (TIM complex) and is assisted by a chaperone system (mitochondrial heat shock protein 70 (mtHsp70) and its partner proteins). Understanding the mechanism of protein folding is not only important for uncovering basic biophysical principles underlying these fundamental processes [131-134], but also critical to developing novel therapeutics for targeting them [142].

Over the last five decades, biophysical ensemble techniques, including fluorescence, circular dichroism and hydrogen exchange measured by nuclear magnetic resonance or mass spectrometry, combined with stopped-flow technology, have been applied to follow protein unfolding and refolding processes in real time [143-145]. In the last twenty years, single-molecule approaches have been developed, these include fluorescence techniques [146-148] such as Förster resonance energy transfer, and force spectroscopy [149-151], notably by using atomic force microscopy, magnetic tweezers and optical tweezers. However, both the ensemble techniques and single-molecule approaches involve two-end pulling and releasing, and therefore cannot contribute to our understanding of cellular protein folding, which, as

described in the preceding text, occurs at the exit of biological channels and usually initiates from one terminus.

In recent studies, nanopore-mediated protein unfolding was investigated using a model biological pore, *Staphylococcal aureus* α-hemolysin (αHL). The small proteins, Smt3 [96,100] and thioredoxin [97,106] from *Escherichia coli*, were each co-translocationally unfolded (Chapter 1, Section 1.3.3 and Section 1.3.4). This was achieved either by using an electric potential acting on an oligonucleotide attached to the protein N or C terminus, or a molecular motor, the unfoldase ClpX. The step-like unfolding process, different from that which occurs by chemical denaturation [107,108], plausibly results from the pore acting as a barrier at its entrance and initiating the protein unfolding in a different pathway.

In the present work, we have built on these initial studies with nanopores to investigate multiple cycles of protein unfolding and refolding. Notably, we present an approach by which the time dependence of refolding of thioredoxin can be observed by repeated measurements on individual molecules. Two biotinylated oligonucleotide handles were covalently attached to a thioredoxin variant at both termini (Chapter 2). The oligonucleotide-attached thioredoxin bearing a single traptavidin bound at one of the two biotinylated oligonucleotide ends unfolded as the unbound oligonucleotide was electrophoretically pulled through an αHL pore under a voltage bias. By binding the other biotin on the oligonucleotide handle in the other side of the pore, a single molecule of thioredoxin could be captured within the pore and

subjected to repeated cycles of translocation and retro-translocation. The refolding process initiated after retro-translocation through the pore in the recipient compartment. The rate of refolding was determined by introducing a variable delay before the re-translocation step, and comparing the re-translocation duration, after each refolding time, with native translocation. In this controlled manner, fundamental information pertinent to the integrated process of biological pore-mediated protein unfolding and refolding was obtained.

4.2 Results

4.2.1 Capture of a single molecule in the nanopore with traptavidin

We carried out analyses of thioredoxin folding kinetics with the conjugate btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn (Chapter 2). Because traptavidin, a streptavidin variant, binds biotin with more than tenfold slower biotin dissociation than the wild-type streptavidin [152], for the long-time scale single-channel recordings we used the monovalent traptavidin and traptavidin to doubly cap the construct's biotinylated oligonucleotide handles at both the *cis* and *trans* sides of an αHL pore. They were expected to work as "brakes" to halt the translocated protein escaping through either side of the pore.

As with the previous experiments (Chapter 3, Section 3.2.3), the conjugates were reacted with approximately half stoichiometric amounts of the monovalent traptavidin (mTA) to yield mixtures: dual-bound constructs (mTA-btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn-mTA), mono-bound constructs (btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn-mTA or mTA-btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn) and non-bound constructs (btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn). These mixtures were added to the *cis* compartment of a chamber, and +100 mV potential was applied. Four kinds of translocation events were observed as described previously (Chapter 3, Figure 3.5 and Figure 3.8).

To capture a single molecule of thioredoxin within the pore, traptavidin was added to the *trans* compartment. For example, with a C terminus-first

translocation of mTA-btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn (Figure 4.2), when monovalent traptavidin was bound to the N-terminal biotinylated oligonucleotide and fixed at the *cis* entrance of the pore, the C-terminal biotinylated oligonucleotide was extruded into the *trans* side (④ in Figure 4.2 b). Excess amounts of traptavidin were then added to the *trans* side to bind the biotin on the C-terminal oligonucleotide (⑤ in Figure 4.2 b). Thus, a single molecule of thioredoxin was doubly captured at both sides of an αHL pore. This model system was then used to investigate protein refolding by repeatedly retro-translocating the single molecule back to the *cis* side, refolding for a given amount of time and retranlocating to the *trans* side again. This will be discussed in detail below.

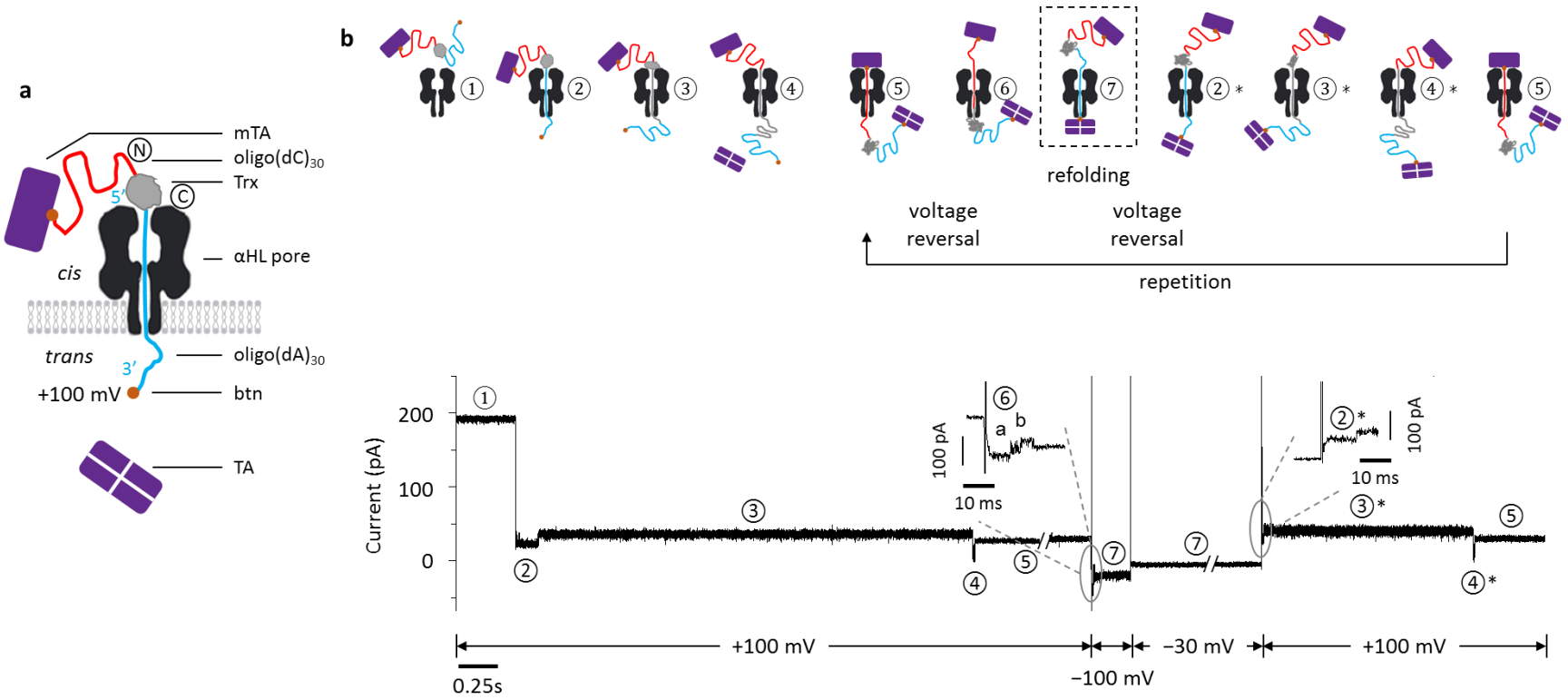


Figure 4.2 I Repeated nanopore (re)translocations of a single thioredoxin stoppered at both ends. (a) Schematics depicting a single molecule of thioredoxin double captured with monovalent traptavidin and traptavidin at both sides of an αHL pore. (b) Upper panel: C terminus-first translocation of the construct (mTA-btn-oligo(dC)$_{30}$-S1C-TrxV5-C109-oligo(dA)$_{30}$-btn) with monovalent traptavidin bound to the N terminus (①→④). The C terminus is captured in the *trans* side with traptavidin (⑤). After this second capture, the thioredoxin can be subjected to cycles of retro-translocation (⑤→⑥), delaying to refold (⑦) and re-translocation (②*→④*). Lower panel: representative ionic current signal corresponding to repeated (re)translocations of a single thioredoxin.

4.2.2 Continuous unfolding and refolding of a single molecule

In practice, for the C terminus-first translocation of the construct, after the double capture of the thioredoxin with monovalent traptavidin and traptavidin within the pore, the potential was reversed to −100 mV, and the protein was pulled retro-translocating to the *cis* side of the pore (⑥ in Figure 4.2 b). The protein was then allowed to refold for various delay times (0.09 s, 0.48 s, 0.98 s, 1.48 s, 10 s, 100 s and 1000 s) (⑦). After this refolding step, +100 mV potential was subsequently applied to drive the refolded protein back through the pore, with it co-retranslocationally unfolding (②*→④*). Using a periodic voltage switching control method (Chapter 5, Section 5.3.3, Figure 5.3), this process was repeated 84–300 times for each above delay time in three independent experiments. A series of re-translocation events with multi-level current blockades for the refolded protein were monitored by single-channel recordings.

Different from the simple four-level current blockade (level 2, 3, 4 and 5, Figure 4.2 b) produced by the first passage of native protein during its translocation through the pore (level 5 represents the oligonucleotide immobilized within the pore, and the protein is held back by the oligonucleotide), those of refolded protein re-translocation were varied. Besides the current signal with all four blockade levels (Figure 4.3 a), other current signals showed only two or three of these levels, especially when the protein refolded for short delay times. Referring to the values of I_{RES}% [97], the levels of these current signals were determined accordingly: current

signals with level 2, 4 and 5 (Figure 4.3 b), with level 3, 4 and 5 (Figure 4.3 c), and with level 4 and 5 only (Figure 4.3 d). In this experiment, the amplified signal was filtered at 5 kHz; transit current blockades with dwell time less than 0.2 ms were filtered, and so undetectable.

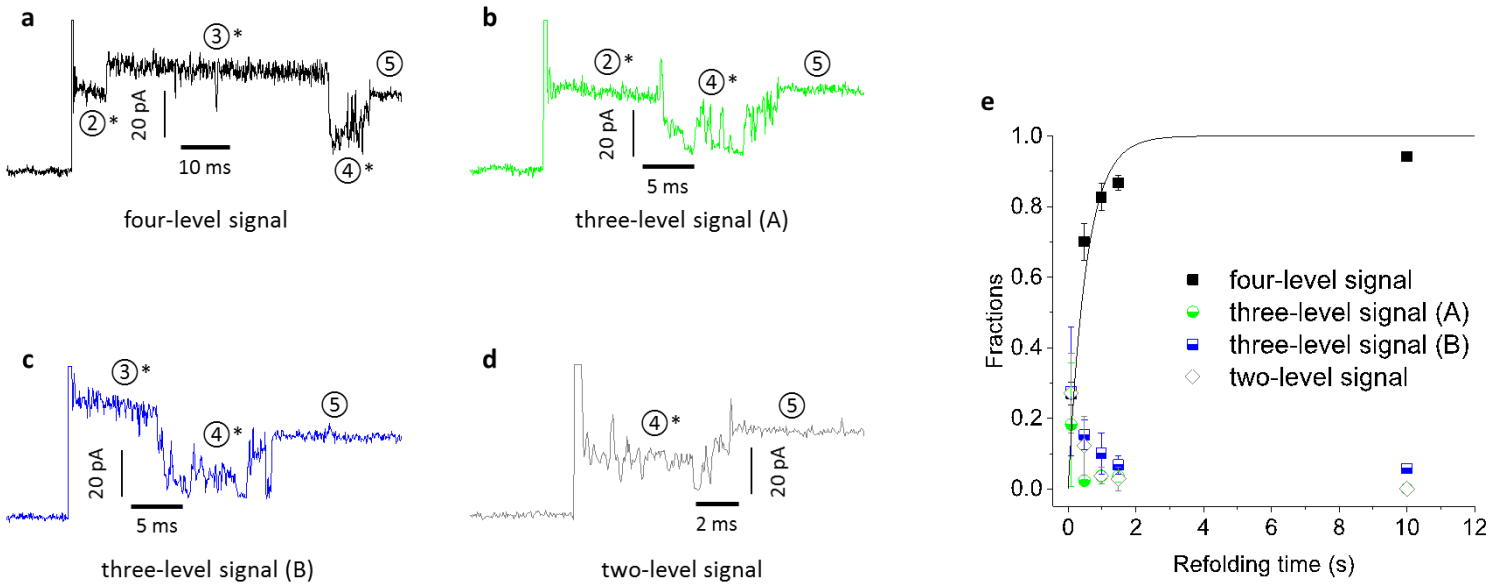


Figure 4.3 I Typical ionic current signals of co-retranslocational unfolding of a refolded thioredoxin. (a-d) Four types of ionic current signals upon re-translocation after refolding in state ⑦. Depending on the refolding time, events with all four levels (level 2, 3, 4 and 5, a), events with three levels (level 2, 4 and 5, b; level 3, 4 and 5, c), events with two levels (level 4 and 5, d) were detected by single-channel recordings. (e) The four types of ionic current signals as fractions of the total number of re-translocations versus the refolding time in state ⑦. The fractions of "four-level signal" as a function of refolding time were fitted with a single exponential yielding a single rate of 1.86 ± 0.86 s^{-1}. Error bars represent standard deviations from three independent experiments. The applied voltage was +100 mV, and the experiments were conducted at 22.9 °C. Current signals produced at other temperatures are shown in Figure 4.10.

As described in the preceding text (Chapter 1, Section 1.3.4), the current blockades of level 2 and 3 represent the native thioredoxin being pulled by the oligonucleotide, and sequentially unfolded and translocated through the pore. Notably, in the experiment of co-translocational unfolding of the chemically denatured protein, the fractions of the current signal lacking both levels 2 and 3 increased as the concentration of denaturing agent increased; nearly half of the current signals showed one blockade level (level 4 only) in 4.8 M urea [97]. The chemically unfolded proteins were pulled passing through the pore freely, and the temporal current fluctuations of level 2 and 3 were not detected by the equipment. Analogous to the chemical denaturation, the disappearances of distinct current level(s) in current work were a result of protein refolding for short delay times. Therefore, the current signal lacking both level 2 and 3 represents a fully unfolded conformation, and the current signal lacking one of the levels represents partially folded conformations. As expected, level 4 appears in the re-translocation of refolded proteins that were refolded for all the delay times, including the very short delay times of 0.09 s and 0.48 s. This confirms that level 4 represents a state where the remaining polypeptide chain of the unfolded protein diffuses through the pore [97] and is thereby not involved in any folded structure. These four types of current signal were plotted as fractions of the total number of re-translocations versus the refolding time (Figure 4.3 e). The fraction of the standard “four-level signal” was fitted with a single-exponential yielding a single rate of 1.86 ± 0.86 s^{-1} at 22.9 °C. It should be noted that the presence of the “four-level signal” did not necessarily represent that the protein was fully refolded – the dwell times in

each level for the co-retranslocational unfolding of refolded protein could be shorter than the co-translocational unfolding of the native one. Quantitative analysis of this case is provided in the following section.

4.2.3 Describing folding kinetics of thioredoxin with a nanopore

The dwell times in current blockade of level 2 and 3 correspond to the duration required to co-translocationally unfold the protein through the pore and reflect the mechanical stability of the local structure presented to the pore [97,106]. To gain further insight into the refolding extent of the protein to its native state, we compared the dwell times in each current level for refolded protein with those for native one (Figure 4.4). In single-channel recordings, a well-established method to analyze the distributions of dwell times is to use constant logarithmic bin widths, by which the exponential components appear as peaked functions in the display [153] (see also references [154-156] for details of single-channel recording data analysis). This representation is especially useful when comparing different histograms where the dwell times vary over a wide range and when fitting histograms containing multiple exponential components.

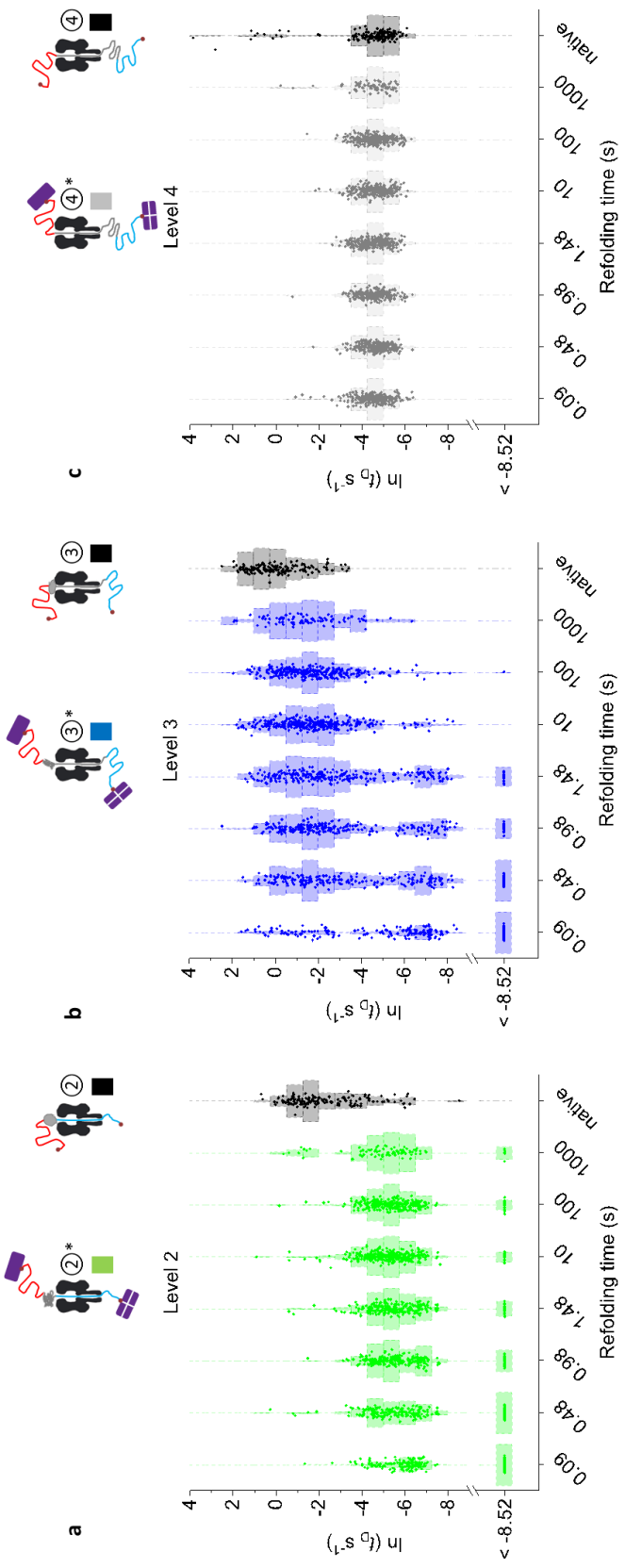


Figure 4.4 I Duration of each current level for co-retranslocational unfolding of refolded thioredoxin. Upper panels: Schematics depicting the states of co-retranslocational unfolding of refolded protein and co-translocational unfolding of native protein, which correspond to the current blockade of (a) level 2, (b) level 3 and (c) level 4. Lower panels: Scatter plots of the logarithmically binned dwell times (ln (f_D s^{-1})) in each level for refolded and native proteins. The refolded proteins were allowed to refold for various delay times (0.09 s, 0.48 s, 0.98 s, 1.48 s, 10 s, 100 s and 1000 s) before co-retranslocational unfolding. Three independent experiments each contributed an equal fraction of the plotted data; in total: 0.09 s (300 events), 0.48 s (255 events), 0.98 s (264 events), 1.48 s (279 events), 10 s (324 events), 100 s (306 events), 1000 s (84 events). The data from individual experiments are in Figure 4.6 (level 2), Figure 4.8 (level 3) and Figure 4.9 (level 4). The applied voltage was +100 mV, and the experiments were conducted at 22.9 °C. Data for experiments conducted at other temperatures are shown in Figure 4.11.

In general, it is apparent that for the refolded protein, refolding for various delay times, shows shorter dwell times in level 2 and 3 than the native one (Figure 4.4 a, b). Those dwell times become longer and approach the native as the refolding times extend long enough. Events that lack distinct current levels (t_D < 0.2 ms, ln (t_D s^{-1}) < −8.52) are also plotted as single bars after breaks in each scatter.

For the quantitative analysis of the translocation duration, the distribution of logarithmically binned dwell time (ln (t_D s^{-1})) in level 2 for the native protein was found to be monomodal and was fitted to the following exponential function (black in Figure 4.4 a and Figure 4.5 a):

$$f(x) = A\frac{1}{t_P}\exp(x - \frac{1}{t_P}\exp(x))$$

where A is the amplitude, and t_P denoting the peak location (most probable dwell time). It yielded a most probable translocation dwell time in level 2, $t_P^{L2, N}$ = 2.57 ± 0.08 × 10^{-1} (Table 4.1) (averaged over three replicates).

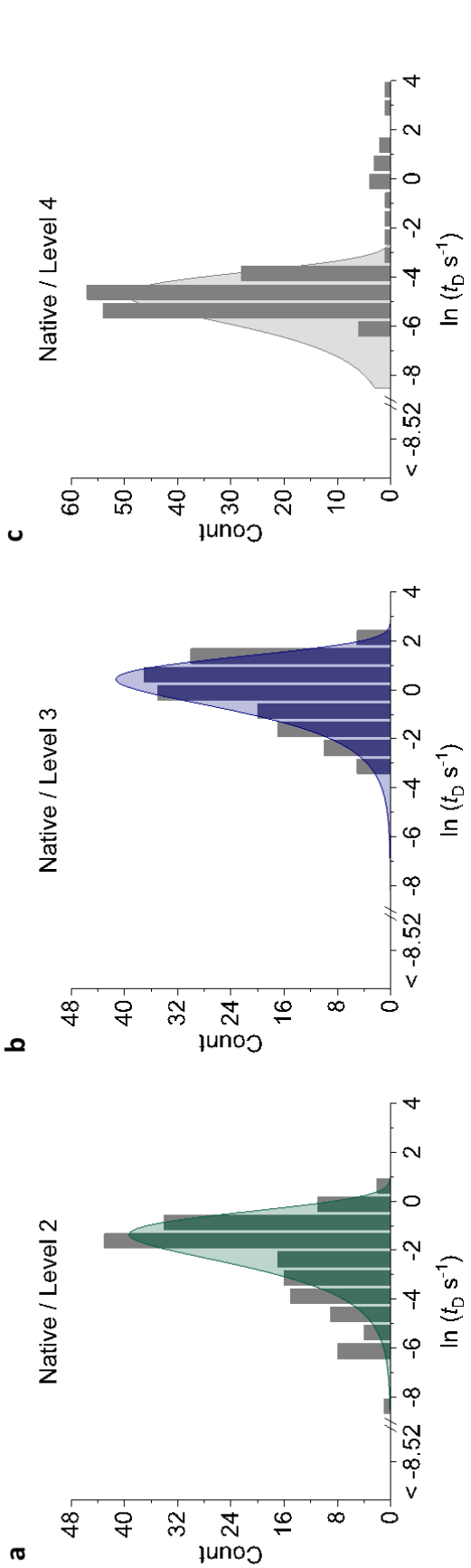


Figure 4.5 I Duration of each current level for co-translocational unfolding of native thioredoxin. (a) Event histogram of ln (t_D s^{-1}) in level 2 for co-translocational unfolding of native thioredoxin. The fit is to a single-exponential function and yields a most probable translocation dwell time, $t_P^{L2, N}$ = 2.57 ± 0.08 × 10^{-1}. (b) Event histogram of ln (t_D s^{-1}) in level 3. The fit is to a single-exponential function and yields $t_P^{L3, N}$ = 1.59 ± 3.21. (c) Event histogram of ln (t_D s^{-1}) in level 4. The fit is to a single-exponential function and yields $t_P^{L4, N}$ = 8.96 ± 1.17 × 10^{-3}. Data were collected from three independent experiments. In total, 160 events were analyzed. The applied voltage was +100 mV, and the experiments were conducted at 22.9 °C.

Table 4.1 l t_P values in each level for co-(re)translocational unfolding of native and refolded thioredoxin

	Experiments	$t_P^{L2, N}$,s		$t_P^{L3, N}$,s		$t_P^{L4, N}$,s
Native	1	2.58×10^{-1}		1.35		7.98×10^{-3}
	2	2.46×10^{-1}		1.37		8.29×10^{-3}
	3	2.66×10^{-1}		2.04		1.06×10^{-2}
	Mean (S.D.)	$2.57 \pm 0.08 \times 10^{-1}$		1.59 ± 3.21		$8.96 \pm 1.17 \times 10^{-3}$
		$t_P^{L2, F}$,s	$t_P^{L2, I}$,s	$t_P^{L3, F}$,s	$t_P^{L3, I}$,s	$t_P^{L4, R}$,s
Refolded	1	2.82×10^{-1}	6.34×10^{-3}	6.76×10^{-1}	1.55×10^{-3}	1.05×10^{-2}
	2	3.18×10^{-1}	8.52×10^{-3}	1.79×10^{-1}	1.20×10^{-3}	1.18×10^{-2}
	3	2.62×10^{-1}	3.79×10^{-3}	3.42×10^{-1}	1.09×10^{-3}	1.41×10^{-2}
	Mean (S.D.)	$2.88 \pm 0.23 \times 10^{-1}$	$6.21 \pm 1.93 \times 10^{-3}$	$3.99 \pm 2.07 \times 10^{-1}$	$1.28 \pm 0.19 \times 10^{-3}$	$1.21 \pm 0.15 \times 10^{-2}$

However, the distributions of ln (t_D s^{-1}) in level 2 for the refolded protein, after various refolding times, were found to be bimodal (green in Figure 4.4 a, Figure 4.6). More variation existed among three replicates than that of the native one, and therefore the three replicates were plotted separately in Figure 4.6. The two exponential components correspond to the long-time blockades (minor fractions, dark green components in Figure 4.6) and the short-time blockades (dominant fractions, light green components in Figure 4.6). The double-exponential fits yield two most probable translocation times: for the long-time blockades, $t_P^{L2, F}$ = 2.88 ± 0.23 × 10^{-1} and for the short ones, $t_P^{L2, I}$ = 6.21 ± 1.93 × 10^{-3} (Table 4.1). The value for the long-time blockade of the refolded protein ($t_P^{L2, F}$) approaches the value of the native ($t_P^{L2, N}$), indicating the full recovery of the mechanical resistance in the refolded protein structural elements. We thus define the dark green components fully folded species, the light green components partially folded species, and events lacking the current level (ln (t_D s^{-1}) < −8.52) unfolded species.

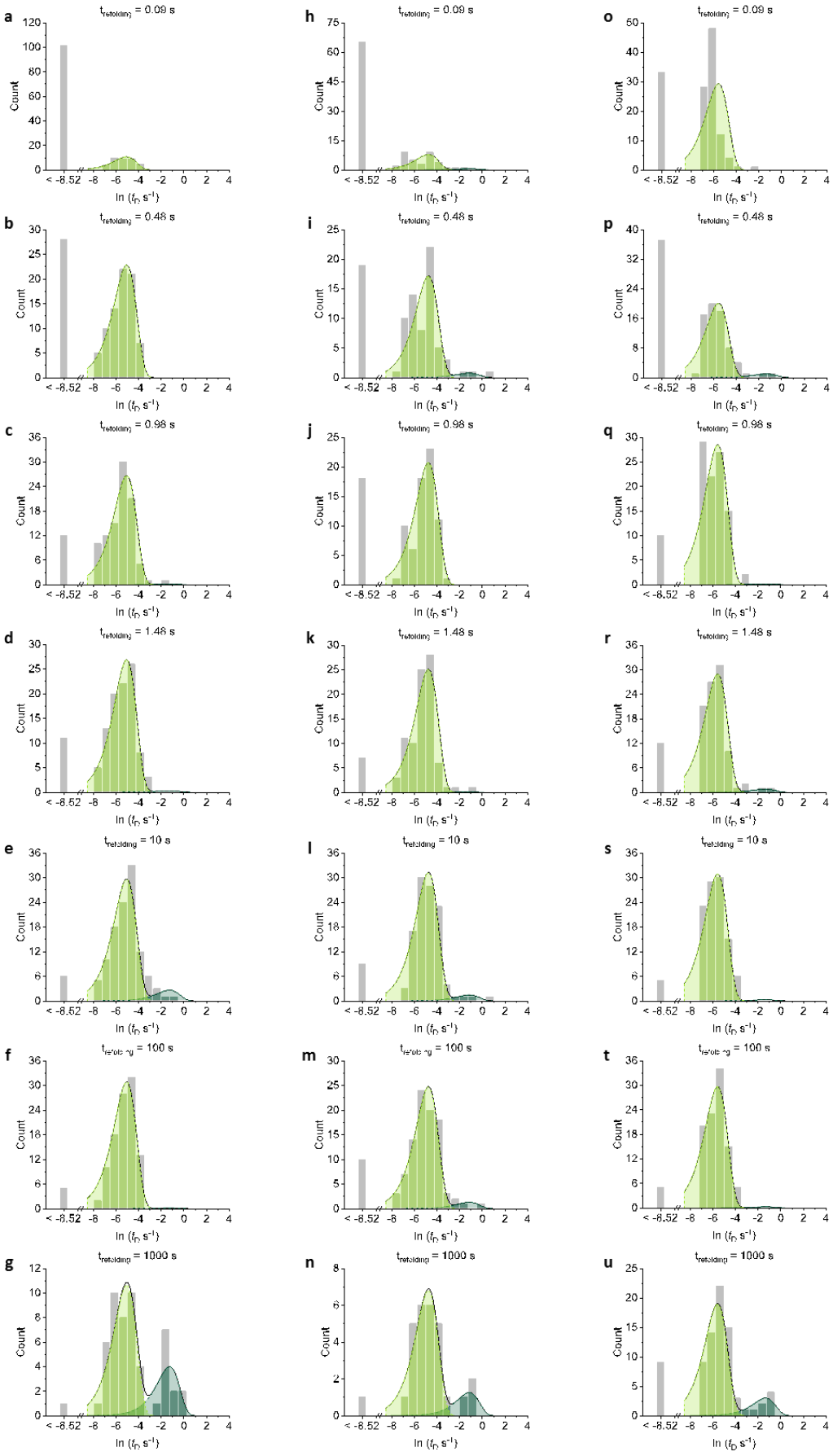

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
Count
ln (k_f, s^{-1})
< -8.52

Figure 4.6 l Duration of current level 2 for co-retranslocational unfolding of refolded thioredoxin in three separate experiments. (a-g) Event histograms of ln (t_D s^{-1}) in level 2 in experiment 1 for co-retranslocational unfolding of refolded thioredoxin, which were refolded for various delay times: (a) 0.09 s, (b) 0.48 s, (c) 0.98 s, (d) 1.48 s, (e) 10 s, (f) 100 s and (g) 1000 s. The fits are to double-exponential functions and yield two t_P^{L2}. For the long-time blockades: $t_P^{L2, F}$ = 2.82 × 10^{-1} (dark green components), and for the short ones: $t_P^{L2, I}$ = 6.34 × 10^{-3} (light green components). The dashed lines are the deconvolution into single exponentials. Events lacking level 2 (ln (t_D s^{-1}) < −8.52) are plotted as single bars after breaks. (h-n) Event histograms in experiment 2. (o-u) Event histograms in experiment 3. The values of $t_P^{L2, F}$ and $t_P^{L2, I}$ in experiment 2 and 3 are listed in Table 4.1. Averaged over these three replicates, $t_P^{L2, F}$ = 2.88 ± 0.23 × 10^{-1} and $t_P^{L2, I}$ = 6.21 ± 1.93 × 10^{-3}. The applied voltage was +100 mV, and the experiments were conducted at 22.9 °C. The numbers of the events analyzed in each refolding time and in each experiment are as follows. Experiment 1: 0.09 s (138 events), 0.48 s (107 events), 0.98 s (107 events), 1.48 s (108 events), 10 s (119 events), 100 s (108 events), 1000 s (52 events); experiment 2: 0.09 s (100 events), 0.48 s (85 events), 0.98 s (88 events), 1.48 s (93 events), 10 s (118 events), 100 s (103 events), 1000 s (28 events); experiment 3: 0.09 s (127 events), 0.48 s (108 events), 0.98 s (105 events), 1.48 s (106 events), 10 s (108 events), 100 s (102 events), 1000 s (80 events). The data from these experiments are combined in Figure 4.4 a (green).

The fraction of refolded protein in each conformation as a function of refolding time is plotted in Figure 4.7 a. Approximately half of the events are in the unfolded conformation (lacking blockade of level 2) when refolding for 0.09 s, and this fraction decreases very rapidly as refolding for slightly longer time. The accumulation to partially folded conformation (short-time blockades of level 2) follows a single-exponential yielding a single rate $k_{U\rightarrow I}^{L2}$ = 2.05 ± 0.65 s^{-1} (light green line in Figure 4.7 a). Further, the accumulation to fully folded conformation (long-time blockades of level 2) is extremely slow with a single rate $k_{U\rightarrow F}^{L2}$ = 1.74 ± 1.14 × 10^{-4} s^{-1} (dark green line in Figure 4.7 a).

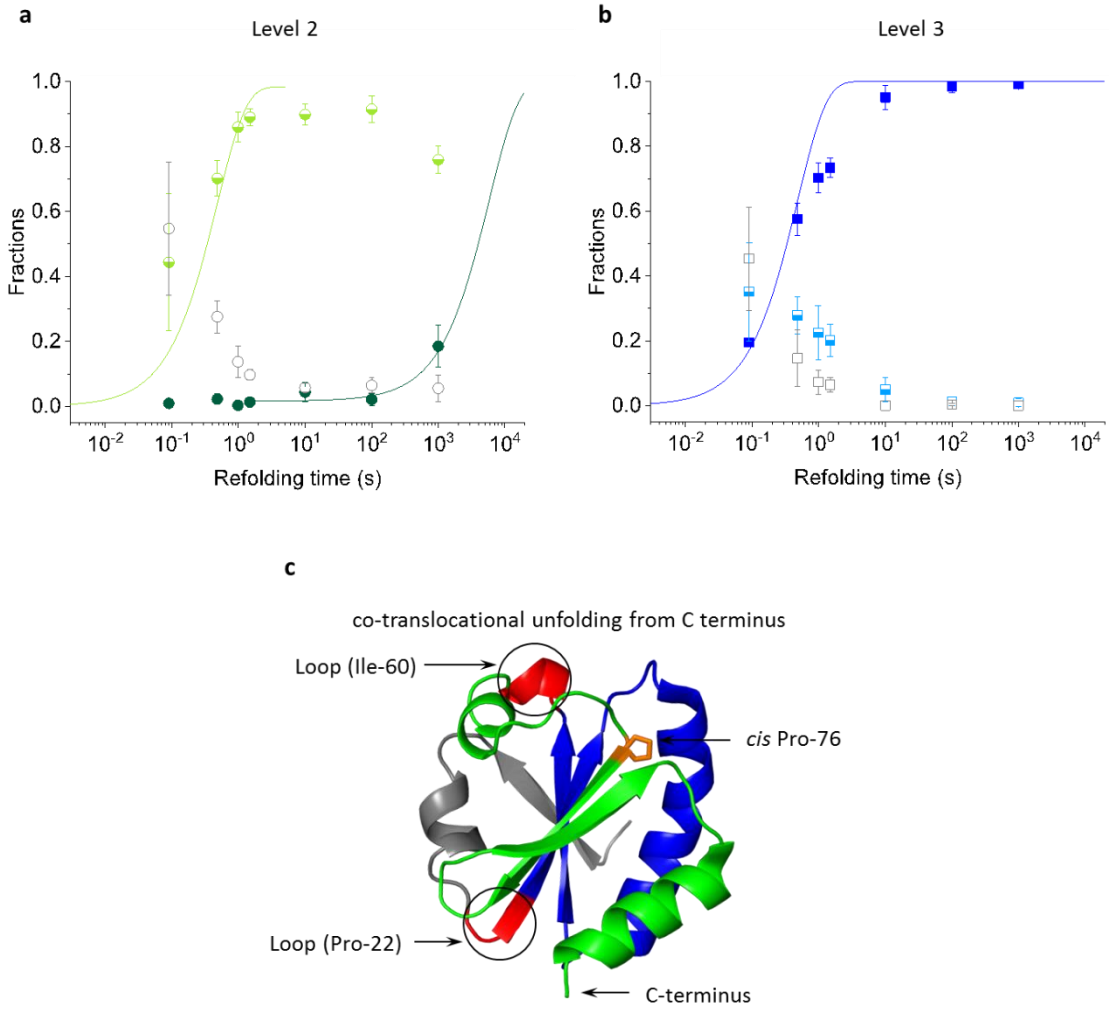


Figure 4.7 I Folding kinetics of thioredoxin by nanopore analysis. (a) Folding kinetics of thioredoxin evaluated by the recovery rates of blockade duration in level 2. The fractions of refolded thioredoxin in fully folded conformation (dark green circles) (dark green components in Figure 4.6), partially folded conformation (light green circles) (light green components in Figure 4.6) and unfolded conformation (grey circles) (single bars after breaks in Figure 4.6) versus the refolding time are plotted. A single-exponential fit yields a single rate $k_{U\rightarrow i}^{L2}$ = 2.05 ± 0.65 s^{-1} for the formation of partially folded conformation (light green line), and $k_{U\rightarrow F}^{L2}$ = 1.74 ± 1.14 × 10^{-4} s^{-1} for the formation of fully folded conformation (dark green line). level 2 represents a state with thioredoxin captured atop the pore, during which the C terminus regions presented to the pore begin to unfold (green segments, c). (b) Folding kinetics of thioredoxin evaluated by the recovery rates of blockade duration in level 3. The fractions of refolded thioredoxin in fully folded conformation (dark blue squares) (dark blue components in Figure 4.8), partially folded conformation (light blue squares) (light blue components in Figure 4.8) and unfolded conformation (grey squares) (single bars after breaks in Figure 4.8) versus

the refolding time are plotted. A single-exponential fit yields a single rate $k_{U\rightarrow I}^{L3}$ = 2.08 ± 0.53 s^{-1} for the formation of fully folded conformation (dark blue line). Level 3 represents a state, after the first unfolding step, with the polypeptide chain within the pore (green segments, c). In this blockade duration, the intermediate (remaining folded regions) begins to unfold (blue segments, c). The grey segments in c are involved in a final diffusion step in the blockade duration of level 4. Error bars represent standard deviations from three independent experiments. The applied voltage was +100 mV, and the kinetics were obtained at 22.9 °C.

The reappearance of the long-time blockades in level 2 as a variable of refolding time ($k_{U\rightarrow F}^{L2}$) is approximately four orders of magnitude slower than that of the short ones ($k_{U\rightarrow I}^{L2}$). We deduced that the folding kinetics evaluated above based on the current blockade of level 2 were involved in a rate-limiting *trans*-to-*cis* isomerization of the Ile75-Pro76 peptide bond: there is an invariant single *cis* Pro-76 buried in native thioredoxin (Figure 4.7 c), and it accumulates a non-native *trans* conformation in the unfolded species [157,158]. In the experiment of co-translocational protein unfolding, level 2 represents a state where the protein is pulled by the oligonucleotide and is captured atop the pore. During this state, the driving force destabilizes the local structure presented to the pore. For the C terminal pulling, the force-bearing portion of thioredoxin involves a segment of ~ 40 amino acids stretching from the C terminus to a loop at Ile-60 (green segment in Figure 4.7 c) [106,159], and the Pro-76 is in this segment. Therefore, the long-time blockades in level 2 were very likely produced during the re-translocation of the refolded thioredoxin with Pro-76 in *cis*, and the short ones were produced by the refolded protein with Pro-76 in *trans*. The protein folds rapidly to an intermediate with *trans* Pro-76, but it reaches the native conformation with *cis*

Pro-76 extremely slowly. These results are in good agreement with the ensemble measurements reported recently [158].

This interpretation became stronger with another experiment, in which co-translocational unfolding kinetics of a thioredoxin variant with no prolines (Trx 0P: S1C/C32S/P34A/C35S/P40A/P64A/P68A/P76A/C109) were investigated (work in progress). The most probable translocation dwell time in level 2 for this native 0P ($t_P^{L2,\,N\,(0P)}$) is 20-fold smaller than that of the native V5 ($t_P^{L2,\,N}$) and the long-time blockades of the refolded V5 ($t_P^{L2,\,F}$) (protein variant in current work), but it approximates the value of the short-time blockades of the refolded V5 ($t_P^{L2,\,I}$), with less than 3-fold larger. It illustrated that the partially folded conformation (short-time blockades) of the refolded V5 exhibited a similar mechanical resistance as the native 0P, implying this partially refolded conformation very likely represents a state that is in a stable 0P-like tertiary structure with the Ile75-Pro76 peptide bond in the nonnative *trans* conformation (like *trans* Ile75-Ala76 peptide bond in 0P).

Similar to the current blockade of level 2, the distribution of ln (t_D s^{-1}) in level 3 for the native protein was also monomodal and was fitted to a single-exponential function, yielding a most probable translocation dwell time in level 3, $t_P^{L3,\,N}$ = 1.59 ± 3.21 (black in Figure 4.4 b, Figure 4.5 b and Table 4.1). Whereas, the distributions for the refolded protein after various refolding times were found to be bimodal (blue in Figure 4.4 b, Figure 4.8): the double-exponential fits yielded $t_P^{L3,\,F}$ = 3.99 ±2.07 × 10^{-1} (long-time blockades, dark blue components in Figure 4.8 and Table 4.1) and $t_P^{L3,\,I}$ = 1.28 ± 0.19 × 10^{-3}

(short-time blockades, light blue components in Figure 4.8 and Table 4.1). Referring to the value of native protein, we attributed the long-time blockades to the re-translocation of fully refolded proteins and the short ones to the re-translocation of partially refolded species. According to level 3, the fraction of refolded protein in each conformation as a function of refolding time was plotted in Figure 4.7 b; the formation rate for the fully folded conformation follows a single-exponential and yields a single rate $k_{U\rightarrow F}^{L3}$ = 2.08 ± 0.53 s^{-1} (dark blue line in Figure 4.7 b). In the unfolding experiment, the duration of level 3 has been proposed to be the time that is required to unfold a segment of native thioredoxin between loops at the residue Ile-60 to Pro-22 (blue segment in Figure 4.7 c). Therefore, in the current work, the obtained rate constant ($k_{U\rightarrow F}^{L3}$) indicates the folding kinetics of the corresponding segment in the unfolded protein.

It should be noted that the value of $t_P^{L3,\,F}$ for the fully refolded species is approximately four times smaller than the value of $t_P^{L3,\,N}$ for the native protein; these two values had been expected to be similar. This could be a result of experimental variation. For example, in experiment 1, the value of $t_P^{L3,\,F}$ is approximately twice smaller than $t_P^{L3,\,N}$. Second, and perhaps more importantly, the nonnative *trans* conformation of the 75-76 peptide bond in the refolded protein, prior to the completion of the exceedingly slow *trans*-to-*cis* isomerization, can translate into conformational changes in the polypeptide segment 30–35 and structural rearrangements in the segment 39–41 [158,160]. These segments in the refolded protein may be destabilized and

will be released from the remaining polypeptide more easily during the co-retranslocational unfolding in level 3.

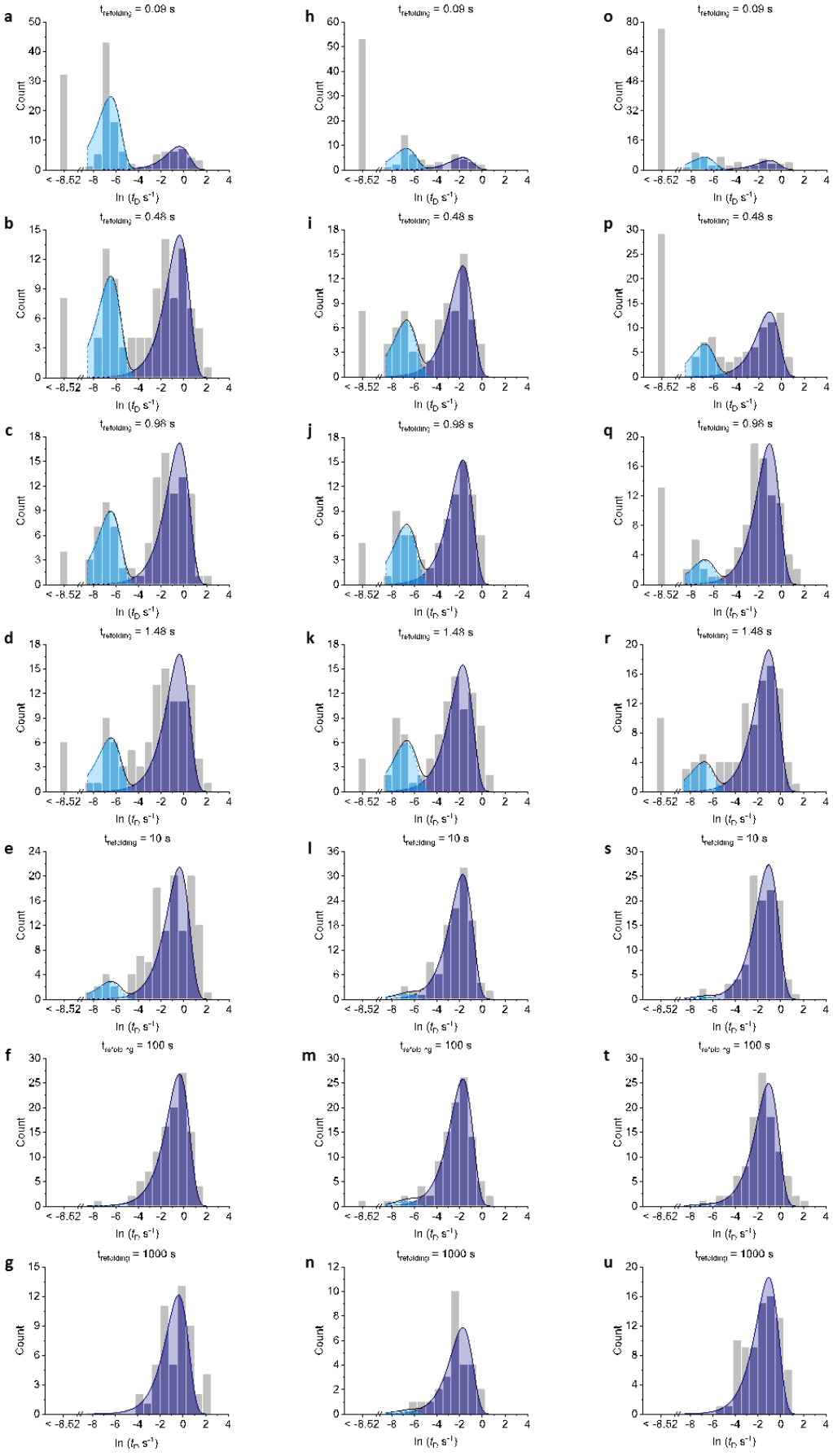

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
Count
ln (t_D s^{-1})
ln (t_n s^{-1})

Figure 4.8 I Duration of current level 3 for co-retranslocational unfolding of refolded thioredoxin in three separate experiments. (a-g) Event histograms of ln (t_D s^{-1}) in level 3 in experiment 1 for co-retranslocational unfolding of refolded thioredoxin, which were refolded for various delay times: (a) 0.09 s, (b) 0.48 s, (c) 0.98 s, (d) 1.48 s, (e) 10 s, (f) 100 s and (g) 1000 s. The fits are to double-exponential functions and yield two t_P^{L3}. For the long-time blockades: $t_P^{L3, F}$ = 6.76 × 10^{-1} (dark blue components), and for the short ones: $t_P^{L3, I}$ = 1.55 × 10^{-3} (light blue components). The dashed lines are the deconvolution into single exponentials. Events lacking level 3 (ln (t_D s^{-1}) < −8.52) are plotted as single bars after breaks. (h-n) Event histograms in experiment 2. (o-u) Event histograms in experiment 3. The values of $t_P^{L3, F}$ and $t_P^{L3, I}$ in experiment 2 and 3 are listed in Table 4.1. Averaged over these three replicates, $t_P^{L3, F}$ = 3.99 ± 2.07 × 10^{-1} and $t_P^{L3, I}$ = 1.28 ± 0.19 × 10^{-3}. The applied voltage was +100 mV, and the experiments were conducted at 22.9 °C. The numbers of the events analyzed in each experiment are listed in the caption of Figure 4.6. The data from these experiments are combined in Figure 4.4 b (blue).

As expected, the ln (t_D s^{-1}) in level 4 for the re-translocation of refolded protein, after various refolding times, showed similar single-exponential distributions. Averaged over the data for all the refolding times in three replicates, the single-exponential fits yielded a most probable translocation dwell time, $t_P^{L4, R}$ = 1.21 ± 0.15 × 10^{-2} (grey in Figure 4.4 c, Figure 4.9 and Table 4.1), and this value approaches the native, $t_P^{L4, N}$ = 8.96 ± 1.17 × 10^{-3} (black in Figure 4.4 c, Figure 4.5 c and Table 4.1). Level 4 occurs after the protein completely unfolds in level 2 and 3. It has been proposed that this level begins with a hairpin-like structure containing Pro-22 and Val-23 clogging the αHL pore constriction (grey segment in Figure 4.7 c). Once it disentangles, the remaining polypeptide chain diffuses through the pore [106,159]. The diffusion of this segment during level 4 is not involved in any folded structure, and therefore the dwell times in level 4 show no dependence on the refolding times.

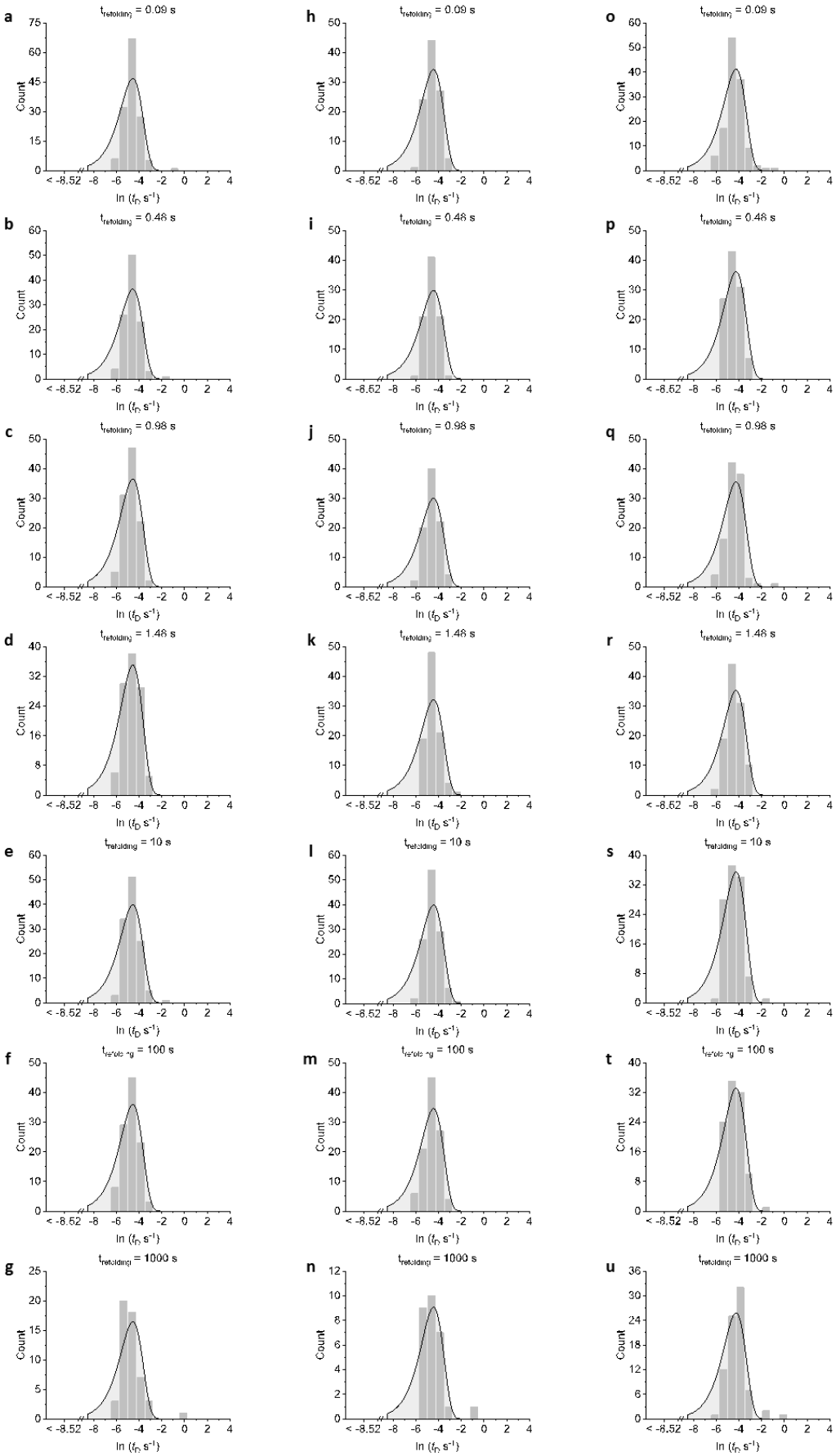

a
$t_{refolding}$ = 0.09 s
b
$t_{refolding}$ = 0.48 s
c
$t_{refolding}$ = 0.98 s
d
$t_{refolding}$ = 1.48 s
e
$t_{refolding}$ = 10 s
f
$t_{refolding}$ = 100 s
g
$t_{refolding}$ = 1000 s
h
$t_{refolding}$ = 0.09 s
i
$t_{refolding}$ = 0.48 s
j
$t_{refolding}$ = 0.98 s
k
$t_{refolding}$ = 1.48 s
l
$t_{refolding}$ = 10 s
m
$t_{refolding}$ = 100 s
n
$t_{refolding}$ = 1000 s
o
$t_{refolding}$ = 0.09 s
p
$t_{refolding}$ = 0.48 s
q
$t_{refolding}$ = 0.98 s
r
$t_{refolding}$ = 1.48 s
s
$t_{refolding}$ = 10 s
t
$t_{refolding}$ = 100 s
u
$t_{refolding}$ = 1000 s
Count
ln (t_D s^{-1})
< -8.52
-8
-6
-4
-2
0
2
4

Figure 4.9 l Duration of current level 4 for co-retranslocational unfolding of refolded thioredoxin in three separate experiments. (a-g) Event histograms of ln (t_D s^{-1}) in level 4 in experiment 1 for co-retranslocational unfolding of refolded thioredoxin, which were refolded for various delay times: (a) 0.09 s, (b) 0.48 s, (c) 0.98 s, (d) 1.48 s, (e) 10 s, (f) 100 s and (g) 1000 s. The fits are to single-exponential functions and yield t_P^{L4} = 1.05 × 10^{-2}. (h-n) Event histograms in experiment 2. (o-u) Event histograms in experiment 3. The values of t_P^{L4} in experiment 2 and 3 are listed in Table 4.1. Averaged over these three replicates, t_P^{L4} = 1.21 ± 0.15 × 10^{-2}. The applied voltage was +100 mV, and the experiments were conducted at 22.9 °C. The numbers of the events analyzed in each experiment are listed in the caption of Figure 4.6. The data from these experiments are combined in Figure 4.4 c (grey).

We also tested whether temperature would have any effect on the refolding and co-retranslocational unfolding of a single protein. We found that the duration of each current level decreases as the temperature increases (current signals in Figure 4.10, distributions of ln (t_D s^{-1}) in Figure 4.11). The decreases of dwell time in level 2 and 3 could be attributed to either heating increasing the co-retranslocational protein unfolding rate or slowing the protein refolding process; the decrease in level 4 suggested that increasing temperature accelerated the diffusion of polypeptide chain through the pore. Moreover, the residue currents (I_{RES}) were observed increased with temperature (Figure 4.10), and this phenomenon was due to the strong temperature dependence of single-channel conductance [91]. Thus, the folding kinetics of thioredoxin described in this work were obtained at a constant temperature of 22.9 °C.

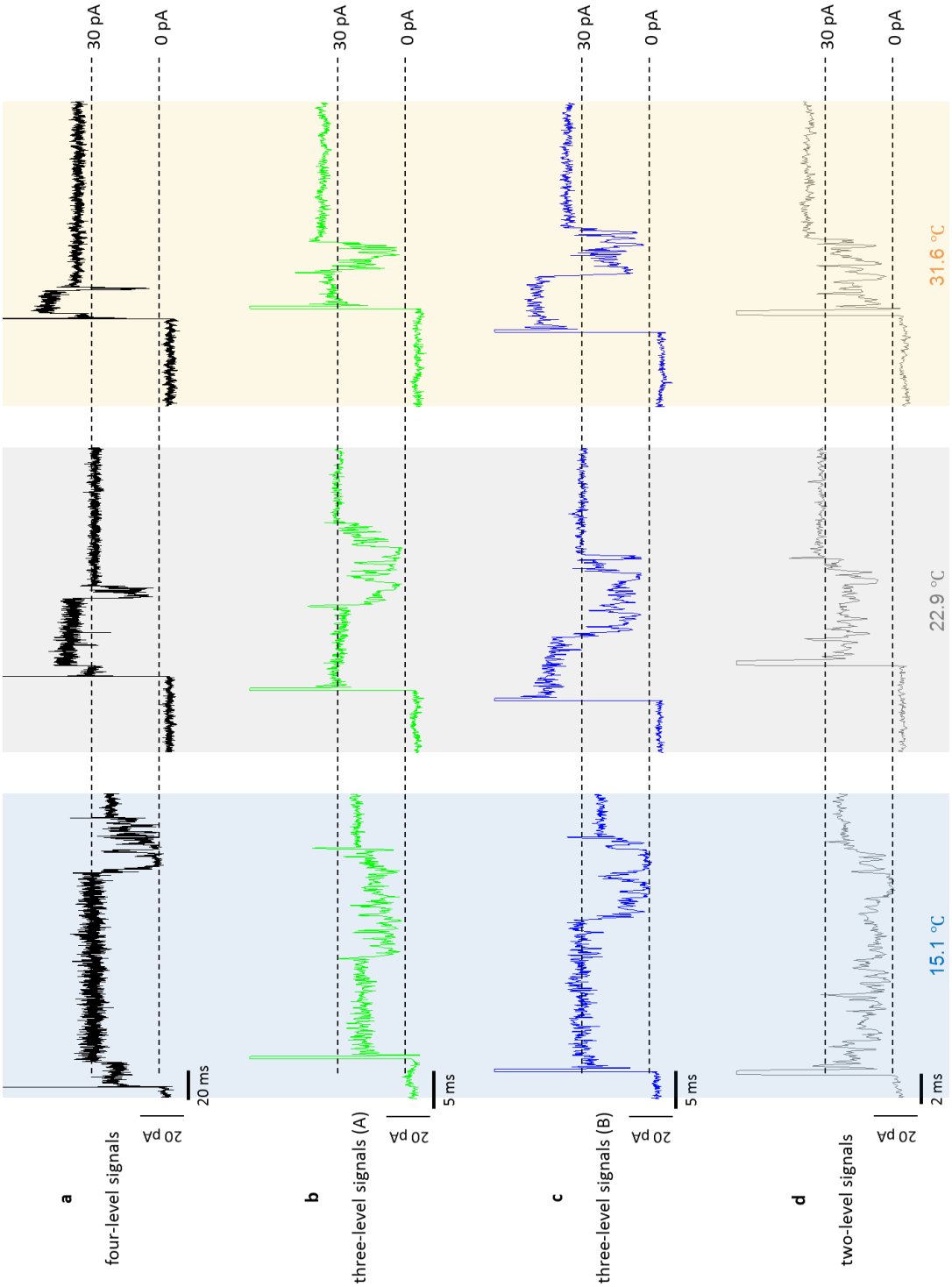


Figure 4.10 I Typical ionic current signals of co-retranslocational unfolding of refolded thioredoxin at different temperatures. The four types of ionic current signals (described in Figure 4.3 a-d) produced at 15.1 °C (blue background), 22.9 °C (grey background) and 31.6°C (yellow background). The applied voltage was +100 mV.

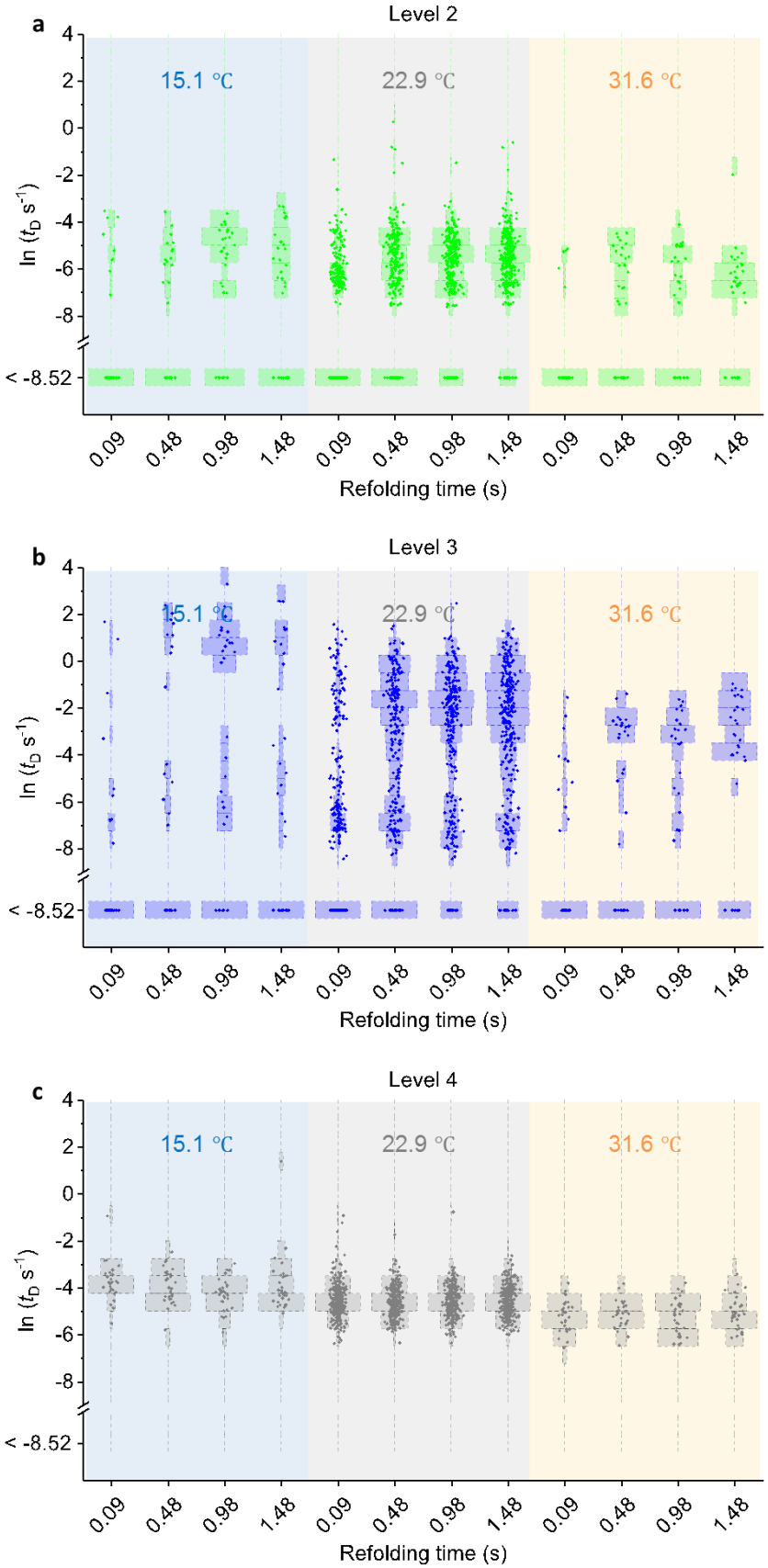

a
Level 2
15.1 ℃
22.9 ℃
31.6 ℃
ln (t_D s^{-1})
4
2
0
-2
-4
-6
-8
< -8.52
0.09
0.48
0.98
1.48
Refolding time (s)
b
Level 3
15.1 ℃
22.9 ℃
31.6 ℃
ln (t_D s^{-1})
Refolding time (s)
c
Level 4
15.1 ℃
22.9 ℃
31.6 ℃
ln (t_D s^{-1})
Refolding time (s)

Figure 4.11 l Duration of each current level for co-retranslocational unfolding of refolded thioredoxin at different temperatures. Scatter plots of ln (t_D s^{-1}) in (a) level 2, (b) level 3 and (c) level 4 for co-retranslocational unfolding of refolded thioredoxin, which were refolded for various delay times (0.09 to 1.48 s). The cycles of retro-translocation, delaying to refold and re-translocation were at 15.1 °C (blue background), 22.9 °C (grey background) and 31.6°C (yellow background). Data were collected from three independent experiments. In total, 30 events were analyzed for each temperature of 15.1 °C and 31.6°C, and for each refolding time; the data for the temperature of 22.9 °C are from Figure 4.4. The applied voltage was +100 mV.

4.3 Discussion

Electrical recording of a single αHL pore has previously been used to detect, at the single-molecule level, proteins in folded, partially folded and unfolded states (Chapter 1, Section 1.3.2) [27,94,95,161], and to investigate co-translocational protein unfolding during its passing through the pore (Chapter 1, Section 1.3.3 and 1.3.4) [96,97,100,106]. Taking advantage of these abilities, we present a new method to study protein folding by the repeated measurement of the unfolding and refolding of an individual protein (thioredoxin), which is captured within the αHL pore.

In the experiment, following the refolding of a single protein, initiated from a retro-translocated unfolded state, the protein was pulled back through the pore with co-retranslocational unfolding. We found that the protein could continually refold and maintain its mechanical integrity for more than 700 cycles of unfolding and refolding, over more than 10 h. Various conformational states that depended on the time allowed for refolding were detected. These conformational states were characterized by the dwell times of the current blockades during the re-translocation and by the comparison with the translocation of native protein (Figure 4.12). The results illustrated that, for the repeated process of unfolding and refolding, an individual unfolded protein could reach the native state through different microscopic folding trajectories. In addition, protein-to-protein and/ or pore-to-pore variations existed, which were reflected by the differences in the most probable dwell times (t_P) for each conformational state among three independent experiments (Table 4.1).

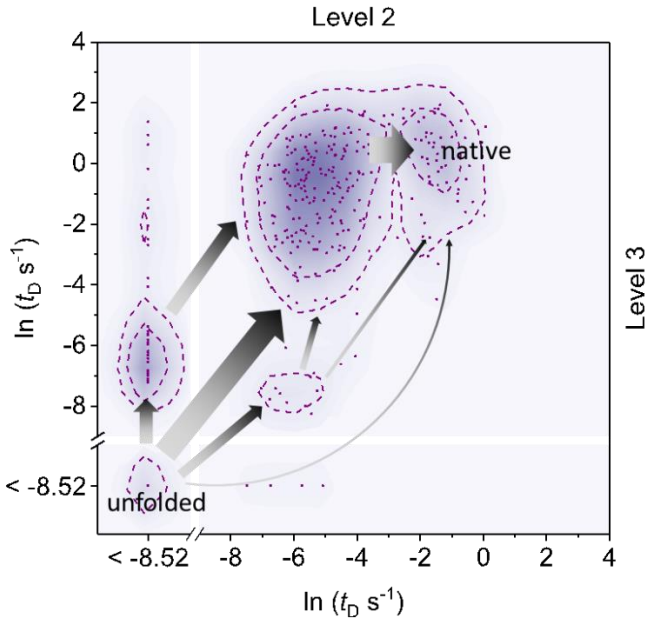


Figure 4.12 I Thioredoxin refolding flux. During the period when an individual unfolded thioredoxin reached the native state, various conformational states with different dwell times in level 2 and 3 were detected (purple dots). Black arrows indicate assumed folding trajectories between each pair of conformations, and the thickness of the arrows reflects the assumed frequencies of the conformational transitions. The data are combined from Figure 4.5 a, b, Figure 4.6 a-g and Figure 4.8 a-g.

The microscopic folding view of thioredoxin presented here is different from the two-state folding description of this model protein deduced from conventional macroscopic experiments [107,108]. The anomalous observations of various conformational states during the folding or unfolding process of a two-state folder have also been reported elsewhere [27,97,162,163]. A first explanation is that non-rate limiting steps that are hidden in bulk measurements could be discerned at single-molecule resolution [164]; short-lived intermediates or metastable partially folded states could be detected by single-molecule techniques [162,163,165]. Second, in

the co-translocational protein unfolding experiments, the pore presents a physical barrier at its entrance, which could force the protein to take a different unfolding pathway with intermediate states [97]. In this work, protein refolding might follow the two-state transition. However, the time dependence of protein refolding was investigated by pulling the refolded protein back through the pore with co-retranslocational unfolding, and therefore various intermediates were detected during this process. Third, in the present work, the folding might begin coincidently with the retro-translocation: the space of the vestibule of αHL pore could allow the translocating polypeptide chain to form small tertiary structure elements (a state between ⑥and ⑦ in Figure 4.2), before it fully escapes through the pore and freely folds in solution at state ⑦ (for the structure of αHL pore, see Chapter 1, Section 1.1.3). It will be interesting to test this interpretation by using shorter oligonucleotide handles (e.g. 15-mer). In this manner, the protein can be held back by the oligonucleotide and confined within the vestibule, and folding may occur inside. If this is the case, this co-translocational folding will be topologically restricted, and not all the elements of the polypeptide chain will be simultaneously available for folding while translocating. Therefore, it differs fundamentally from mechanical two-end pulling and releasing during force spectroscopy measurements or refolding of a chemically or thermally denatured protein *in vitro*.

The model system established here offers several advantages over previous approaches for protein folding research and may have more significant implications. Following the capture of a single protein within the pore,

perfusion of the recording compartments allows an effectively infinite dilution of the excess protein (see Methods in Chapter 5, Section 5.3.3), which prevents the formation of aggregates during the refolding. The protein studied here, thioredoxin, was separated from the stopper proteins, monovalent traptavidin and traptavidin, by oligonucleotide handles, thereby reducing the possibility of nonspecific protein-protein interactions. Additionally, the cycles of repeated translocation might be used to reduce reading errors during the nanopore analysis of post-translational modifications [111] and, potentially, nanopore proteomics. Finally, and more importantly, this system platform may provide new insights into protein folding process *in vivo*, which usually occurs at the exit of biological channels. In this aspect, future work might include examining the effects of chaperone-mediated protein folding.

CHAPTER 5

Methods

5.1 General methods

Unless otherwise noted, the chemicals and solvents used were of analytical grade and were used as received from commercial sources. Distilled, deionized water (ddH_2O), used as reaction solvent or buffer, was deionized using a Milli-Q Synthesis system supplied with a Q-Gard 1 Purification Cartridge (Merck Millipore). Analytical thin-layer chromatography (TLC) was performed on TLC Silica gel 60 F_{254} plates (Merck Millipore) with visualization by UV irradiation at 254 nm. Plates were also stained with ninhydrin solution (3 mg/mL, in 95% ethanol) and then dried with a heat-gun to visualize amines

and amides. Oligonucleotide and protein concentrations were determined using a ND-1000 Nano spectrophotometer (Thermo Scientific).

5.2 Bioconjugation of oligonucleotide to protein

5.2.1 Instrumentation and sample analysis

Oligonucleotide mass spectra were measured on a Bruker microTOF spectrometer using an ACQUITY UPLC Oligo BEH C18, 130Å, 1.7 µm, 2.1 × 100 mm Column. Oligonucleotide UV absorbance was monitored at 260 and 280 nm. The mass spectra acquisition was carried out in negative ion mode, and the total oligonucleotide masses were obtained by deconvoluting multiply charged oligonucleotide mass spectra using MaxEnt 1. The mobile phase consisted of 200 mM hexafluoroisopropanol (HFIP), 8.2 mM triethylamine (TEA) buffer in ddH_2O as buffer A, and 20% buffer A in 80% methanol (v/v) as buffer B, at a flow rate of 0.3 mL min^{-1}.

Protein mass spectra were measured on a Waters LCT Premier XE Mass Spectrometer equipped with electrospray interface. The column is a Merck Chromolith C18 2 × 5 mm guard column. Calibration was performed in positive ion mode using horse heart myoglobin. The mass spectra acquisition was carried out in positive ion mode, and the total protein masses were obtained by deconvoluting multiply charged protein mass spectra using Masslynx 4.1 with openLynx open access. The mobile phase consisted of 0.1 % formic acid in ddH_2O as buffer A, and acetonitrile as buffer B, at a flow rate of 0.4 mL min^{-1} (initial, with 95% buffer A and 5% buffer B), 0.4 mL min^{-1} (1 min, with 95% buffer A and 5% buffer B), 0.4 mL min^{-1} (5 min, with 100% buffer B), 1 mL min^{-1} (8 min, with 100% buffer B), 1 mL min^{-1} (8.1 min, with 95% buffer A and 5% buffer B).

^{1}H NMR spectra were measured with a Bruker AVIII HD 400 (400 MHz) spectrometer. ^{1}H NMR chemical shifts are reported as δ in units of parts per million (ppm) relative to dimethyl sulfoxide-d6 (DMSO-d6) (δ 2.50, quintet). Multiplicities are reported as follows: s (singlet), d (doublet), dd (doublet of doublets). Coupling constants are reported as a *J* value in Hertz (Hz). The number of protons (n) for a given resonance is indicated as nH and is based on spectral integration values.

HPLC was performed on an Agilent 1260 Infinity HPLC System. Signal for all HPLC experiments was achieved with an in-line diode array detector. Agilent Polaris 5 C18-A 150 × 4.6 mm and Supelcosil PLC-18 12 µm, 250 × 21.2 mm columns were used for analysis and purification respectively. The mobile phase consisted of 0.1 % formic acid in ddH_2O as buffer A and 0.1 % formic acid in acetonitrile as buffer B, at a flow rate of 1 mL min^{-1} for analysis and 5 mL min^{-1} for purification.

FPLC was performed on ÄKTApurifier. Size-exclusion chromatography was performed with a Superdex 75 10/300 column (GE Healthcare). Ion exchange chromatography was performed with a HiTrap Q HP column or HiTrap Q FF column (GE Healthcare) eluted with 0 to 1 M KCl in 10 mM TrisHCl, 1 mM EDTA, pH 8.0.

SDS-PAGE analysis was performed on a Criterion Cell (Bio-Rad) using an 18% Criterion TGX Precast Midi Protein Gel (Bio-Rad) according to manufactures protocol. The samples and electrode buffers were prepared according to Laemmli. Gels were run for 50 min at 150 V to separate the

bands. Commercially available markers (Bio-Rad) were applied to at least one lane of each gel for assignment of apparent molecular masses. Visualization of protein bands was accomplished by staining with InstantBlue Protein Stain (Expedeon). Gel imaging was performed on a Gel Doc XR+ Gel Documentation System (Bio-Rad).

5.2.2 Synthesis of crosslinker

Crosslinker, **1**. To 1 mL of anhydrous DMSO was added 6-amino-2-cyanobenzothiazole (CBT) (17.5 mg, 0.1 mmol) and *p*-maleimidophenyl isocyanate (PMPI) (10.7 mg, 0.05 mmol). After stirring at room temperature overnight, the pure product was obtained by prep-HPLC. The solvents were removed under reduced pressure by rotary evaporation followed by freeze drying.

5.2.3 Oligonucleotide modification

A 5'-thiol (hexamethylene linker)- and 3'-biotin (triethylene glycol spacer)-modified oligo$(dC)_{30}$ was obtained from Biomers and dissolved in 10 mM TrisHCl, 1 mM EDTA, pH 8.0. The oligonucleotide was reduced with 5 mM TCEP at 65 °C for 1.5 h and then the excess TCEP was removed using an Amicon Ultra column (0.5 mL, 3 kDa, SigmaAldrich), according to the manufacturer's instructions. The oligonucleotide was then treated with 15 equiv of **1** (dissolved in a minimum volume of DMSO) at room temperature for

2 h. The reaction was monitored by HPLC. The modified oligonucleotide (btn-oligo$(dC)_{30}$-**1**) was purified by ion exchange chromatography.

A 5'-thiol (hexamethylene linker)- and 3'-biotin (triethylene glycol spacer)-modified oligo$(dA)_{30}$ was reduced as described above. The oligonucleotide was then activated with 20 mM of 2,2'-dipyridyl disulphide (dissolved in a minimum volume of acetonitrile) at room temperature for 2 h [166]. The modified oligonucleotide (btn-oligo$(dA)_{30}$-**2**) was purified by ion exchange chromatography.

5.2.4 Protein expression and purification

The Trx V5-S1C/C109 gene was cloned into the pET 30a+ plasmid (TopGene) [97]. The mutant used in this work was made with Quick Change II XL site-directed mutagenesis kit (Stratagene) and verified by DNA sequencing. The plasmid was transformed into *Escherichia coli* BL21(DE3) cells (Novagen), and thioredoxin was produced by isopropyl-β-D-1-thiogalactopyranoside induction during the exponential growth phase in 0.25 L of LB medium. The expressed thioredoxin was purified by DNA precipitation with streptomycin sulphate, followed by size-exclusion chromatography and ion exchange chromatography.

To remove pyruvate from the N-terminal cysteine of Trx V5-S1C/C109, it was rebuffered in the deprotection buffer (100 mM sodium phosphate, 150 mM NaCl, 400 mM methoxyamine, 5 mM TCEP, pH 7.0) by Amicon Ultra (0.5 mL,

3 kDa) and incubated at room temperature overnight [113]. The excess methoxyamine was removed, and thioredoxin was rebuffered in 10 mM TrisHCl, 1 mM EDTA, pH 8.0 by using PD MiniTrap G-25 column.

5.2.5 Bioconjugation of oligonucleotide to protein

Excess thioredoxin was mixed with btn-oligo-$(dC)_{30}$-**1** in 10 mM TrisHCl, 1 mM EDTA, pH 8.0 at room temperature overnight [112,113]. The mono-attached conjugate (btn-oligo$(dC)_{30}$-S1C-TrxV5-C109) was purified by ion exchange chromatography. The purified conjugate was immediately mixed with excess btn-oligo$(dA)_{30}$-**2**. The reaction was allowed to proceed overnight at room temperature [166]. The desired dual-attached conjugate (btn-oligo$(dC)_{30}$-S1C-TrxV5-C109-oligo$(dA)_{30}$-btn) was purified by ion exchange chromatography and verified by mass spectrometry.

5.3 Single-channel recordings

5.3.1 αHL pore

The nanopores used in this work were wild-type αHL pores. αHL monomers were produced by expression in an *Escherichia coli in vitro* transcription / translation system and oligomerized into αHL heptamers by incubation with rabbit red blood cell membranes [45,167]. The heptameric αHL pores were purified by SDS-polyacrylamide gel electrophoresis [45,167,168].

5.3.2 Chamber

A vertical planar lipid bilayer chamber was used for the experiment (Figure 5.1). The chamber contains two compartments of 500 µL each, designated *cis* and *trans*. The planar lipid bilayer is formed across an orifice (diameter ≈ 100 µm) in a 25 µm thick polytetrafluoroethylene (PTFE) film (Goodfellow) that separates the two compartments [169]. The bigger cylindrical holes (diameter = 5 mm), which are connected to the main compartments, are used for holding electrodes. The smaller holes (diameter = 1 mm) are used for holding microbore PTFE tubing (internal diameter = 560 µm, outside diameter = 1 mm) (Cole-Parmer), which are connected to a perfusion equipment. Each compartment was covered with a lid to limit the evaporation of water over the course of the experiment.

The chamber, embedded in a copper holder with circulating water, is thermalized using a Circulating Bath / Kryostat ET150 - BC20 (Thermo

Fisher). The single-channel recordings were initiated after the stabilization of the temperature. The temperature (22.9 °C) was measured in the electrolyte solution by using a calibrated thermometer.

The assembly of chamber, electrodes and copper holder was contained within a Faraday cage (~50 × 35 × 25 cm^3) to shield the sensitive acquisition electronics from external radiative electrical noise.

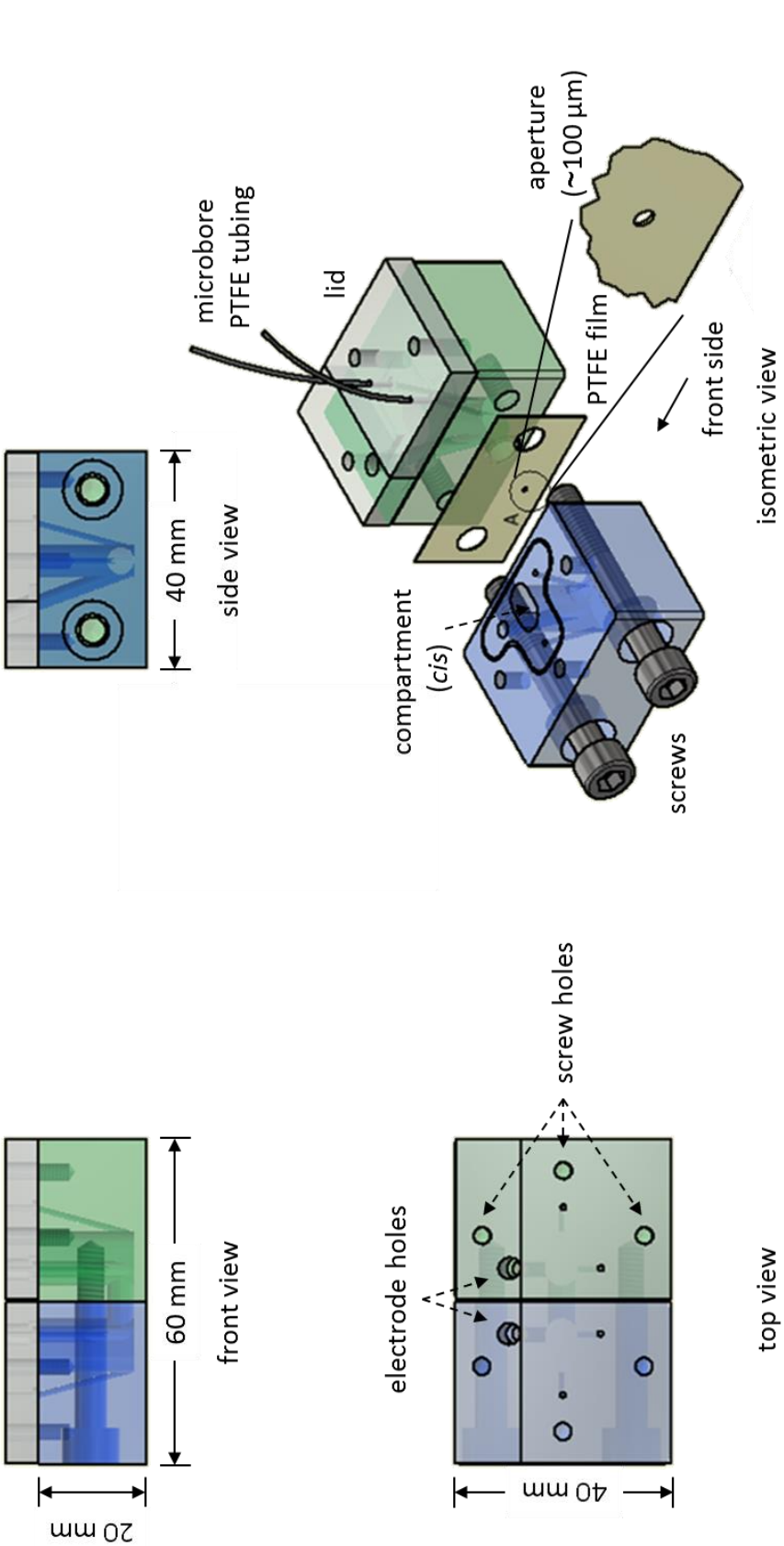


Figure 5.1 I Vertical planar bilayer chamber. The two halves of the chamber are fastened together with screws, clamping the PTFE film between them. The apparatus is sealed with silicone glue to ensure water-tightness. The PTFE film contains a central orifice, across which the bilayer is formed.

5.3.3 Single-channel recordings

The single-channel recordings were carried out by using planar lipid bilayer of 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (DPhPC) (Avanti Polar Lipids) formed across the orifice in the PTFE film of the chamber, as described previously (Figure 5.1) [168,170-172]. To form the bilayer, the orifice was pretreated with hexadecane in *n*-pentane (10% v/v). Electrolyte solution (10 mM HEPES, 2 M KCl, pH 7.2) was added to both compartments, and then the DPhPC in *n*-pentane (10 mg mL^{-1}) was added. The solvent was allowed to evaporate, and the bilayer was formed by lowering and raising the electrolyte level past the orifice.

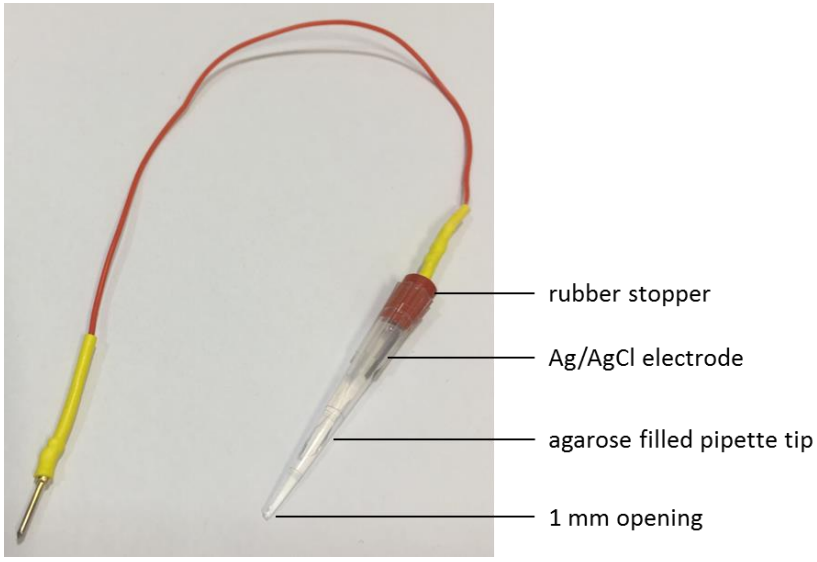


Figure 5.2 I Photograph of an Ag/AgCl electrode. The electrode is incorporated into a 200 mL pipette tip filled with agarose, and the top is sealed with a rubber stopper.

Ag/AgCl electrode was prepared by hypochlorite treatment of a short piece of Ag wire (diameter = 1.5 mm, length ≈10 mm) and then soldered onto a

standard electrical wire. The electrode was incorporated into a 200 mL pipette tip filled with agarose (Bio-Rad) (3%, w/v, 3 M KCl in ddH_2O), and the top was sealed with a rubber stopper (Figure 5.2).

A potential was applied across the bilayer with two electrodes. A positive potential indicates a higher potential in the *trans* compartment, and a positive current is one in which cations flow from the *trans* to the *cis* side. The capacitance of the bilayer was checked and the stability of the bilayer, judged by the absence of fluctuations in the current, was monitored for 1 min at +140 mV.

The αHL pores (~1 ng μL^{-1}, 0.2 µL) were then added to the *cis* compartment, which was held at ground. According to the above convention, the cap domain of the pore [24] is exposed to the *cis* compartment, while the entrance to the transmembrane β-barrel at the tip of the stem domain is exposed to the *trans* side. Following the insertion of a single αHL pore, to prevent further insertions, the *cis* compartment was perfused to remove excess αHL by using a push-pull syringe driver (PHD 2000 Syringe Pump, Harvard Apparatus), which exchanged the electrolyte solution at the rate of 1 mL min^{-1}.

The single-channel currents were amplified by using a patch-clamp amplifier (Axopatch 200B, Axon Instruments), filtered with a low-pass Bessel filter (80 dB/decade) with a corner frequency of 5 kHz, and then digitized with a Digidata 1440A converter (Axon Instruments) at a sampling frequency of 20 kHz. The acquisition software was Clampex 10.6 (Molecular Devices). In the experiment of measuring protein folding kinetics, a periodic voltage switching

program was used for the repeated cycles of retro-translocation, delaying to refold and re-translocation (Figure 5.3). Each experimental condition was repeated at least three times in this work.

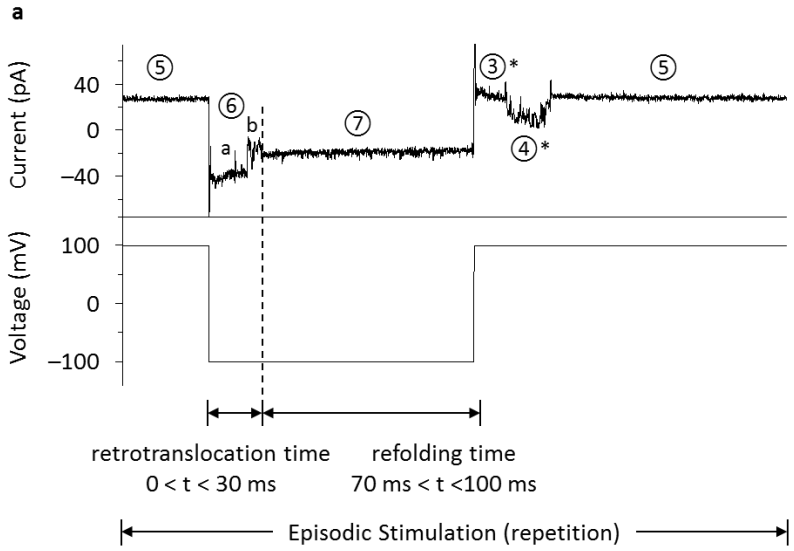
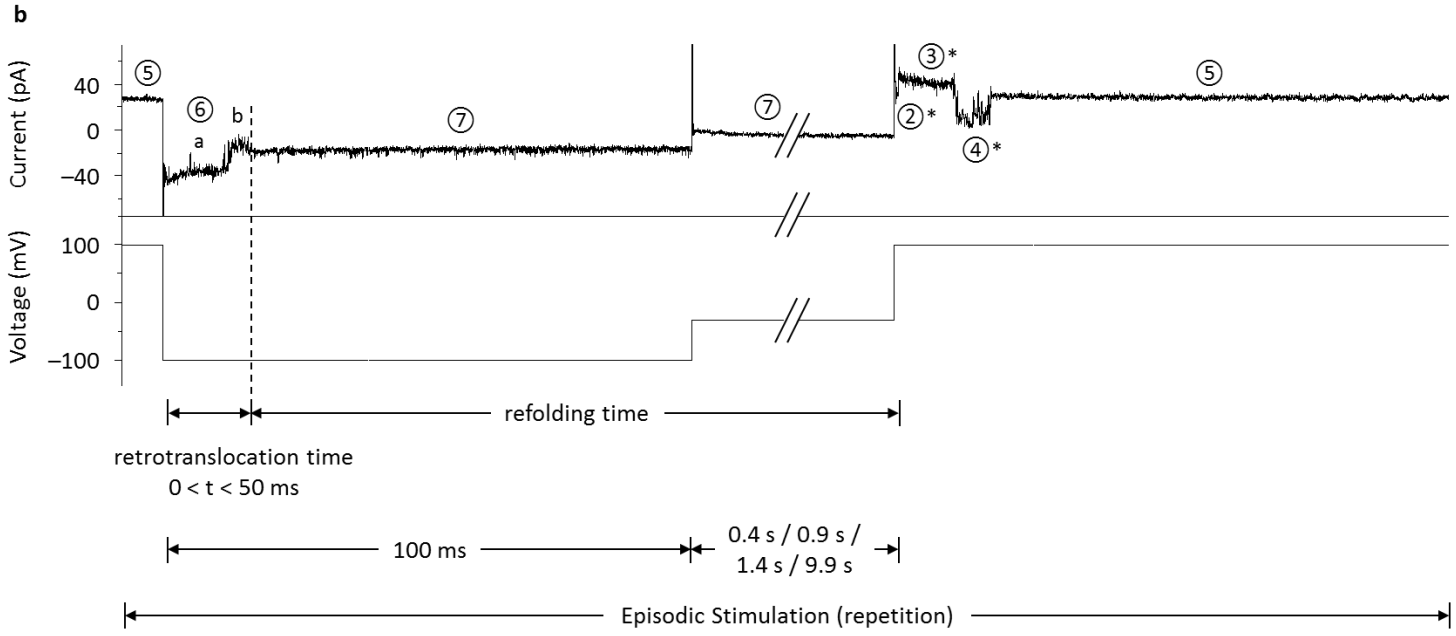
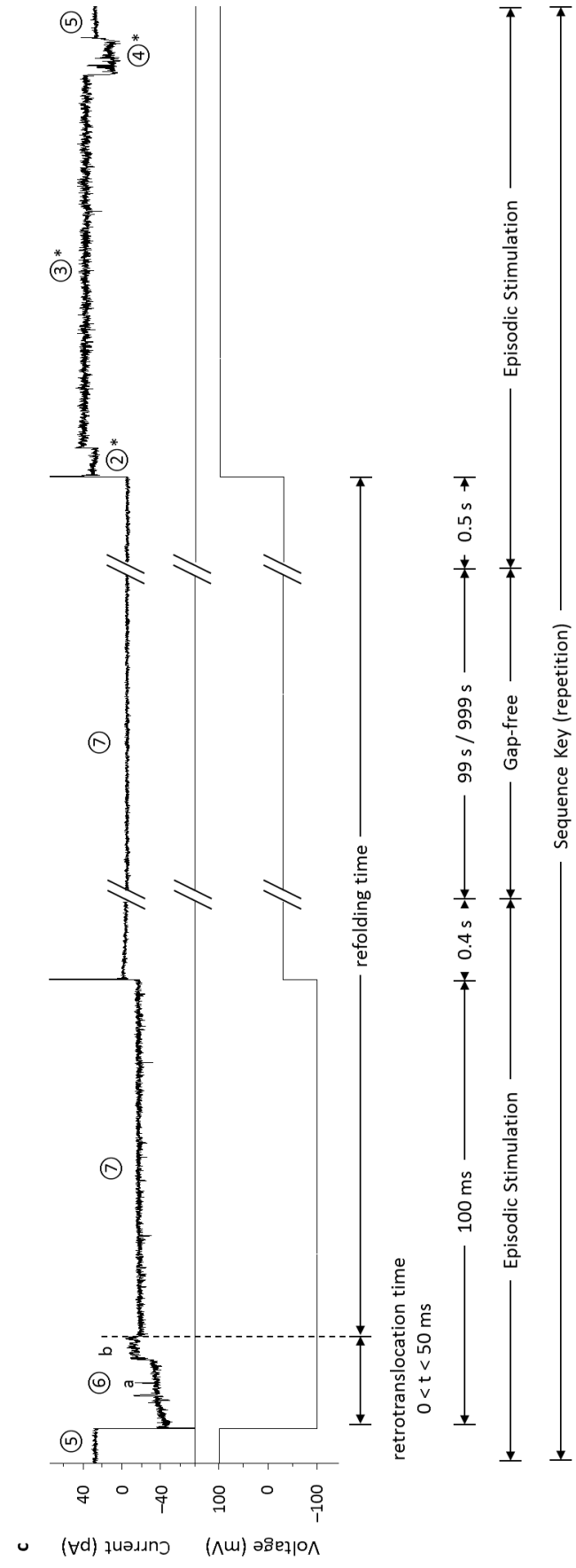


Figure 5.3 I Periodic voltage switching control for repeated cycles of retro-translocation, delaying to refold and re-translocation. (a-b) For the short-time scale experiments (0.1–10 s), the periodic voltage switching was manipulated by an Episodic Stimulation program of the data acquisition software (Clampex 10.6). The refolding time (⑦) in (a) corresponds to the delay time in the applied negative voltage (100 ms at ‒100 mV) deducting the retro-translocation times (⑥). At –100 mV, ~ 70% events showed the duration of retro-translocation less than 30 ms; events with the longer duration were discarded. Therefore, the refolding times for the selected events were 70–100 ms and were averaged to be 0.09 s (see Figure 4.4). (continued on next page)



(continued from previous page) The refolding time (⑦) in (b) corresponds to the delay time in the applied negative voltage (100 ms at –100 mV, and 0.4 s, 0.9 s, 1.4 s or 9.9 s respectively at –30 mV) deducting the retro-translocation times (⑥). The applied voltage of –100 mV was switched to −30 mV to avoid the biotin-traptavidin dissociation during the delay times for refolding. At –100 mV, ~ 85% events showed the duration of retro-translocation less than 50 ms; events with the longer duration were discarded. Therefore, the refolding times for the selected events were 0.45–0.50 s, 0.95–1.00 s, 1.45–1.50 s and 9.95–10.00 s, and were averaged to be 0.48 s, 0.98 s, 1.48 s and 9.98 s (10 s) respectively (see Figure 4.4). (continued on next page)



(continued from previous page) (c) For the long-time scale experiments (100–1000 s), the periodic voltage switching was manipulated by Episodic Stimulation and Gap-free program integrated by a Sequence Key in the data acquisition software (Clampex 10.6). The retro-translocation times were ignored, and the refolding times were 100 s and 1000 s respectively (see Figure 4.4).

5.3.4 Data analysis

Data were analyzed and presented with Clampfit software 10.6 (Molecular Devices) and OriginPro 2017 (OriginLab). Residual current (I_{RES}%) value for each level (level 2 and level 5) was determined as $I_{RES}\% = I_B / I_O \times 100\%$, where I_B is the mean current blockade in each level and I_O is the mean current through the open pore. Current blockades (level 2) below the resolution time (< 0.2 ms) were discarded in the residual current and noise analysis; in the event duration analysis, they were plotted as single bars ($\ln (t_D\ s^{-1}) < -8.52$) after the break in the scatters.

Conclusion

Previous studies with model systems have demonstrated the ability of nanopores, including the αHL pore, to discriminate among different nucleotides, to distinguish among folded, partly unfolded and unfolded states of proteins and to measure co-translocational protein unfolding rates. Here, we have taken advantage of this capability to both determine the translocation direction (N or C terminus-first) of a model protein (thioredoxin) as it unfolds through the pore, and define the folding landscape after the translocation

In total, we tagged the protein at both N and C termini with different distinguishable biotinylated oligonucleotides: the N terminus conjugation was carried out with a bio-orthogonal cyanobenzothiazole condensation, and the C terminus conjugation by disulfide bond formation. When these dual-tagged proteins translocated bidirectionally through the pore, different current blockade levels were produced by the two different oligonucleotides; analysis

of the blockades revealed the direction of protein translocation. Next, by using traptavidin stoppers, a single molecule of protein was captured within the pore and subjected to repeated cycles of unfolding, translocation and refolding. The results indicate a folding pathway where various conformational states are present, and it differs from the two-state mechanism revealed in bulk solution. An important facet of the present work by comparison with denaturation experiments in solution or other single-molecule methods investigating protein folding and unfolding is that the protein emerges vectorially from the pore, either N or C terminus-first, as it would from a transmembrane channel or ribosome. The geometrical constraints within the entrance of the pore could impact the protein folding and unfolding pathway.

Several improvements can be envisaged that build upon the model system established here. For example, chaperone proteins could be added to either side of the pore to examine the chaperone-mediated protein folding. In addition, the cycles of repeated translocation might be used to reduce reading errors during the nanopore-based protein identification. Finally, the efficient and mild labelling of N-terminal cysteine allows selective terminal modification of proteins enriched from transfected cells. This approach will be useful in the nanopore-based analysis of the modified proteins taken from cells.

Bibliography

1. Bayley, H. & Cremer, P. S. Stochastic sensors inspired by biology. *Nature* **413**, 226-230 (2001).

2. Kasianowicz, J. J., Robertson, J. W., Chan, E. R., Reiner, J. E. & Stanford, V. M. Nanoscopic porous sensors. *Annu Rev Anal Chem (Palo Alto Calif)* **1**, 737-766 (2008).

3. Griffiths, J. The realm of the nanopore. Interest in nanoscale research has skyrocketed, and the humble pore has become a king. *Anal Chem* **80**, 23-27 (2008).

4. Howorka, S. & Siwy, Z. Nanopore analytics: sensing of single molecules. *Chem Soc Rev* **38**, 2360-2384 (2009).

5. Shi, W., Friedman, A. K. & Baker, L. A. Nanopore Sensing. *Anal Chem* **89**, 157-188 (2017).

6. Coulter, W. H. Means for counting particles suspended in a fluid. US patent 2656508 (1953).

7. Tobkes, N., Wallace, B. A. & Bayley, H. Secondary structure and assembly mechanism of an oligomeric channel protein. *Biochemistry* **24**, 1915-1920 (1985).

8. Salisbury, I. G., Timsit, R. S., Berger, S. D. & Humphreys, C. J. Nanometer Scale Electron-Beam Lithography in Inorganic Materials. *Applied Physics Letters* **45**, 1289-1291 (1984).

9. Li, J., Stein, D., McMullan, C., Branton, D., Aziz, M. J. & Golovchenko, J. A. Ion-beam sculpting at nanometre length scales. *Nature* **412**, 166-169 (2001).

10. Ying, Y. L., Cao, C. & Long, Y. T. Single molecule analysis by biological nanopore sensors. *Analyst* **139**, 3826-3835 (2014).

11. Wang, S., Zhao, Z., Haque, F. & Guo, P. Engineering of protein nanopores for sequencing, chemical or protein sensing and disease diagnosis. *Curr Opin Biotechnol* **51**, 80-89 (2017).

12. Bayley, H. & Jayasinghe, L. Functional engineered channels and pores (Review). *Mol Membr Biol* **21**, 209-220 (2004).

13. Dekker, C. Solid-state nanopores. *Nat Nanotechnol* **2**, 209-215 (2007).

14. Liu, Y., Dong, X. & Chen, P. Biological and chemical sensors based on graphene materials. *Chem Soc Rev* **41**, 2283-2307 (2012).

15. Hall, A. R., Scott, A., Rotem, D., Mehta, K. K., Bayley, H. & Dekker, C. Hybrid pore formation by directed insertion of alpha-haemolysin into solid-state nanopores. *Nat Nanotechnol* **5**, 874-877 (2010).

16. Gornall, J. L., Mahendran, K. R., Pambos, O. J., Steinbock, L. J., Otto, O., Chimerel, C. *et al.* Simple reconstitution of protein pores in nano lipid bilayers. *Nano Lett* **11**, 3334-3340 (2011).

17. Bell, N. A., Engst, C. R., Ablay, M., Divitini, G., Ducati, C., Liedl, T. *et al.* DNA origami nanopores. *Nano Lett* **12**, 512-517 (2012).

18. Wei, R., Martin, T. G., Rant, U. & Dietz, H. DNA origami gatekeepers for solid-state nanopores. *Angew Chem Int Ed Engl* **51**, 4864-4867 (2012).

19. Vu, T., Davidson, S. L., Borgesi, J., Maksudul, M., Jeon, T. J. & Shim, J. Piecing together the puzzle: nanopore technology in detection and quantification of cancer biomarkers. *Rsc Advances* **7**, 42653-42666 (2017).

20. Healy, K., Schiedt, B. & Morrison, A. P. Solid-state nanopore technologies for nanopore-based DNA analysis. *Nanomedicine* **2**, 875-897 (2007).

21. Miles, B. N., Ivanov, A. P., Wilson, K. A., Dogan, F., Japrung, D. & Edel, J. B. Single molecule sensing with solid-state nanopores: novel materials, methods, and applications. *Chem Soc Rev* **42**, 15-28 (2013).

22. Arjmandi-Tash, H., Belyaeva, L. A. & Schneider, G. F. Single molecule detection with graphene and other two-dimensional materials: nanopores and beyond. *Chem Soc Rev* **45**, 476-493 (2016).

23. Gouaux, J. E., Braha, O., Hobaugh, M. R., Song, L., Cheley, S., Shustak, C. *et al.* Subunit stoichiometry of staphylococcal alpha-hemolysin in crystals and on membranes: a heptameric transmembrane pore. *Proc Natl Acad Sci U S A* **91**, 12828-12831 (1994).

24. Song, L., Hobaugh, M. R., Shustak, C., Cheley, S., Bayley, H. & Gouaux, J. E. Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **274**, 1859-1866 (1996).

25. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci U S A* **93**, 13770-13773 (1996).

26. Maglia, G., Henricus, M., Wyss, R., Li, Q., Cheley, S. & Bayley, H. DNA strands from denatured duplexes are translocated through engineered protein nanopores at alkaline pH. *Nano Lett* **9**, 3831-3836 (2009).

27. Oukhaled, G., Mathe, J., Biance, A. L., Bacri, L., Betton, J. M., Lairez, D. *et al.* Unfolding of proteins and long transient conformations detected by single nanopore recording. *Phys Rev Lett* **98**, 158101 (2007).

28. Church, G., Deamer, D. W., Branton, D., Baldarelli, R. & Kasianowicz, J. Characterization of individual polymer molecules based on monomer-interface interactions US patent 5795782 (1998).

29. Deamer, D. W. & Akeson, M. Nanopores and nucleic acids: prospects for ultrarapid sequencing. *Trends Biotechnol* **18**, 147-151 (2000).

30. Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T. *et al.* The potential and challenges of nanopore sequencing. *Nat Biotechnol* **26**, 1146-1153 (2008).

31. Venkatesan, B. M. & Bashir, R. Nanopore sensors for nucleic acid analysis. *Nat Nanotechnol* **6**, 615-624 (2011).

32. Schneider, G. F. & Dekker, C. DNA sequencing with nanopores. *Nat Biotechnol* **30**, 326-328 (2012).

33. Bayley, H. Nanopore sequencing: from imagination to reality. *Clin Chem* **61**, 25-31 (2015).

34. Deamer, D., Akeson, M. & Branton, D. Three decades of nanopore sequencing. *Nat Biotechnol* **34**, 518-524 (2016).

35. Akeson, M., Branton, D., Kasianowicz, J. J., Brandin, E. & Deamer, D. W. Microsecond time-scale discrimination among polycytidylic acid, polyadenylic

acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. *Biophys J* **77**, 3227-3233 (1999).

36. Meller, A. & Branton, D. Single molecule measurements of DNA transport through a nanopore. *Electrophoresis* **23**, 2583-2591 (2002).

37. Meller, A., Nivon, L., Brandin, E., Golovchenko, J. & Branton, D. Rapid nanopore discrimination between single polynucleotide molecules. *Proc Natl Acad Sci U S A* **97**, 1079-1084 (2000).

38. Wang, H., Dunning, J. E., Huang, A. P., Nyamwanda, J. A. & Branton, D. DNA heterogeneity and phosphorylation unveiled by single-molecule electrophoresis. *Proc Natl Acad Sci U S A* **101**, 13472-13477 (2004).

39. Mathe, J., Aksimentiev, A., Nelson, D. R., Schulten, K. & Meller, A. Orientation discrimination of single-stranded DNA inside the alpha-hemolysin membrane channel. *Proc Natl Acad Sci U S A* **102**, 12377-12382 (2005).

40. Butler, T. Z., Gundlach, J. H. & Troll, M. A. Determination of RNA orientation during translocation through a biological nanopore. *Biophys J* **90**, 190-199 (2006).

41. Deamer, D. W. & Branton, D. Characterization of nucleic acids by nanopore analysis. *Acc Chem Res* **35**, 817-825 (2002).

42. Henrickson, S. E., Misakian, M., Robertson, B. & Kasianowicz, J. J. Driven DNA transport into an asymmetric nanometer-scale pore. *Phys Rev Lett* **85**, 3057-3060 (2000).

43. Kawano, R., Schibel, A. E. P., Cauley, C. & White, H. S. Controlling the Translocation of Single-Stranded DNA through alpha-Hemolysin Ion Channels Using Viscosity. *Langmuir* **25**, 1233-1237 (2009).

44. de Zoysa, R. S., Jayawardhana, D. A., Zhao, Q., Wang, D., Armstrong, D. W. & Guan, X. Slowing DNA translocation through nanopores using a solution containing organic salts. *J Phys Chem B* **113**, 13332-13336 (2009).

45. Maglia, G., Restrepo, M. R., Mikhailova, E. & Bayley, H. Enhanced translocation of single DNA molecules through alpha-hemolysin nanopores by

manipulation of internal charge. *Proc Natl Acad Sci U S A* **105**, 19720-19725 (2008).

46. Rincon-Restrepo, M., Mikhailova, E., Bayley, H. & Maglia, G. Controlled translocation of individual DNA molecules through protein nanopores with engineered molecular brakes. *Nano Lett* **11**, 746-750 (2011).

47. Mitchell, N. & Howorka, S. Chemical tags facilitate the sensing of individual DNA strands with nanopores. *Angew Chem Int Ed Engl* **47**, 5565-5568 (2008).

48. Astier, Y., Braha, O. & Bayley, H. Toward single molecule DNA sequencing: direct identification of ribonucleoside and deoxyribonucleoside 5'-monophosphates by using an engineered protein nanopore equipped with a molecular adapter. *J Am Chem Soc* **128**, 1705-1710 (2006).

49. Wu, H. C., Astier, Y., Maglia, G., Mikhailova, E. & Bayley, H. Protein nanopores with covalently attached molecular adapters. *J Am Chem Soc* **129**, 16142-16148 (2007).

50. Clarke, J., Wu, H. C., Jayasinghe, L., Patel, A., Reid, S. & Bayley, H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol* **4**, 265-270 (2009).

51. Reiner, J. E., Balijepalli, A., Robertson, J. W., Drown, B. S., Burden, D. L. & Kasianowicz, J. J. The effects of diffusion on an exonuclease/nanopore-based DNA sequencing engine. *J Chem Phys* **137**, 214903 (2012).

52. Purnell, R. F., Mehta, K. K. & Schmidt, J. J. Nucleotide identification and orientation discrimination of DNA homopolymers immobilized in a protein nanopore. *Nano Lett* **8**, 3029-3034 (2008).

53. Stoddart, D., Heron, A. J., Mikhailova, E., Maglia, G. & Bayley, H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc Natl Acad Sci U S A* **106**, 7702-7707 (2009).

54. Purnell, R. F. & Schmidt, J. J. Discrimination of single base substitutions in a DNA strand immobilized in a biological nanopore. *ACS Nano* **3**, 2533-2538 (2009).

55. Stoddart, D., Heron, A. J., Klingelhoefer, J., Mikhailova, E., Maglia, G. & Bayley, H. Nucleobase recognition in ssDNA at the central constriction of the alpha-hemolysin pore. *Nano Lett* **10**, 3633-3637 (2010).

56. Stoddart, D., Maglia, G., Mikhailova, E., Heron, A. J. & Bayley, H. Multiple base-recognition sites in a biological nanopore: two heads are better than one. *Angew Chem Int Ed Engl* **49**, 556-559 (2010).

57. Meller, A., Nivon, L. & Branton, D. Voltage-driven DNA translocations through a nanopore. *Phys Rev Lett* **86**, 3435-3438 (2001).

58. Ayub, M., Stoddart, D. & Bayley, H. Nucleobase Recognition by Truncated alpha-Hemolysin Pores. *ACS Nano* **9**, 7895-7903 (2015).

59. Wallace, E. V., Stoddart, D., Heron, A. J., Mikhailova, E., Maglia, G., Donohoe, T. J. *et al.* Identification of epigenetic DNA modifications with a protein nanopore. *Chem Commun (Camb)* **46**, 8195-8197 (2010).

60. Vercoutere, W., Winters-Hilt, S., Olsen, H., Deamer, D., Haussler, D. & Akeson, M. Rapid discrimination among individual DNA hairpin molecules at single-nucleotide resolution using an ion channel. *Nat Biotechnol* **19**, 248-252 (2001).

61. Ashkenasy, N., Sanchez-Quesada, J., Bayley, H. & Ghadiri, M. R. Recognizing a single base in an individual DNA strand: a step toward DNA sequencing in nanopores. *Angew Chem Int Ed Engl* **44**, 1401-1404 (2005).

62. Faller, M., Niederweis, M. & Schulz, G. E. The structure of a mycobacterial outer-membrane channel. *Science* **303**, 1189-1192 (2004).

63. Derrington, I. M., Butler, T. Z., Collins, M. D., Manrao, E., Pavlenok, M., Niederweis, M. *et al.* Nanopore DNA sequencing with MspA. *Proc Natl Acad Sci U S A* **107**, 16060-16065 (2010).

64. Manrao, E. A., Derrington, I. M., Pavlenok, M., Niederweis, M. & Gundlach, J. H. Nucleotide discrimination with DNA immobilized in the MspA nanopore. *PLoS One* **6**, e25723 (2011).

65. Howarth, M., Chinnapen, D. J., Gerrow, K., Dorrestein, P. C., Grandy, M. R., Kelleher, N. L. *et al.* A monovalent streptavidin with a single femtomolar biotin binding site. *Nat Methods* **3**, 267-273 (2006).

66. Benner, S., Chen, R. J., Wilson, N. A., Abu-Shumays, R., Hurt, N., Lieberman, K. R. *et al.* Sequence-specific detection of individual DNA polymerase complexes in real time using a nanopore. *Nat Nanotechnol* **2**, 718-724 (2007).

67. Olasagasti, F., Lieberman, K. R., Benner, S., Cherf, G. M., Dahl, J. M., Deamer, D. W. *et al.* Replication of individual DNA molecules under electronic control using a protein nanopore. *Nat Nanotechnol* **5**, 798-806 (2010).

68. Cockroft, S. L., Chu, J., Amorin, M. & Ghadiri, M. R. A single-molecule nanopore device detects DNA polymerase activity with single-nucleotide resolution. *J Am Chem Soc* **130**, 818-820 (2008).

69. Chu, J., Gonzalez-Lopez, M., Cockroft, S. L., Amorin, M. & Ghadiri, M. R. Real-time monitoring of DNA polymerase function and stepwise single-nucleotide DNA strand translocation through a protein nanopore. *Angew Chem Int Ed Engl* **49**, 10106-10109 (2010).

70. Lieberman, K. R., Cherf, G. M., Doody, M. J., Olasagasti, F., Kolodji, Y. & Akeson, M. Processive replication of single DNA molecules in a nanopore catalyzed by phi29 DNA polymerase. *J Am Chem Soc* **132**, 17961-17972 (2010).

71. Lieberman, K. R., Dahl, J. M., Mai, A. H., Akeson, M. & Wang, H. Dynamics of the translocation step measured in individual DNA polymerase complexes. *J Am Chem Soc* **134**, 18816-18823 (2012).

72. Cherf, G. M., Lieberman, K. R., Rashid, H., Lam, C. E., Karplus, K. & Akeson, M. Automated forward and reverse ratcheting of DNA in a nanopore at 5-A precision. *Nat Biotechnol* **30**, 344-348 (2012).

73. Gyarfas, B., Olasagasti, F., Benner, S., Garalde, D., Lieberman, K. R. & Akeson, M. Mapping the position of DNA polymerase-bound DNA templates in a nanopore at 5 A resolution. *ACS Nano* **3**, 1457-1466 (2009).

74. Manrao, E. A., Derrington, I. M., Laszlo, A. H., Langford, K. W., Hopper, M. K., Gillgren, N. *et al.* Reading DNA at single-nucleotide resolution with a mutant

MspA nanopore and phi29 DNA polymerase. *Nat Biotechnol* **30**, 349-353 (2012).

75. Laszlo, A. H., Derrington, I. M., Brinkerhoff, H., Langford, K. W., Nova, I. C., Samson, J. M. *et al.* Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc Natl Acad Sci U S A* **110**, 18904-18909 (2013).

76. Schreiber, J., Wescoe, Z. L., Abu-Shumays, R., Vivian, J. T., Baatar, B., Karplus, K. *et al.* Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along individual DNA strands. *Proc Natl Acad Sci U S A* **110**, 18910-18915 (2013).

77. Laszlo, A. H., Derrington, I. M., Ross, B. C., Brinkerhoff, H., Adey, A., Nova, I. C. *et al.* Decoding long nanopore sequencing reads of natural DNA. *Nat Biotechnol* **32**, 829-833 (2014).

78. Wescoe, Z. L., Schreiber, J. & Akeson, M. Nanopores discriminate among five C5-cytosine variants in DNA. *J Am Chem Soc* **136**, 16582-16587 (2014).

79. Deamer, D. Nanopore analysis of nucleic acids bound to exonucleases and polymerases. *Annu Rev Biophys* **39**, 79-90 (2010).

80. Steinbock, L. J. & Radenovic, A. The emergence of nanopores in next-generation sequencing. *Nanotechnology* **26**, 074003 (2015).

81. Brown, C. G. & Clarke, J. Nanopore development at Oxford Nanopore. *Nat Biotechnol* **34**, 810-811 (2016).

82. Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A. *et al.* Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol* **36**, 338-345 (2018).

83. Movileanu, L. Interrogating single proteins through nanopores: challenges and opportunities. *Trends Biotechnol* **27**, 333-341 (2009).

84. Oukhaled, A., Bacri, L., Pastoriza-Gallego, M., Betton, J. M. & Pelta, J. Sensing proteins through nanopores: fundamental to applications. *ACS Chem Biol* **7**, 1935-1949 (2012).

85. Sutherland, T. C., Long, Y. T., Stefureac, R. I., Bediako-Amoa, I., Kraatz, H. B. & Lee, J. S. Structure of peptides investigated by nanopore analysis. *Nano Letters* **4**, 1273-1277 (2004).

86. Movileanu, L., Schmittschmitt, J. P., Scholtz, J. M. & Bayley, H. Interactions of peptides with a protein pore. *Biophys J* **89**, 1030-1045 (2005).

87. Wolfe, A. J., Mohammad, M. M., Cheley, S., Bayley, H. & Movileanu, L. Catalyzing the translocation of polypeptides through attractive interactions. *J Am Chem Soc* **129**, 14034-14041 (2007).

88. Parker, M. W., Buckley, J. T., Postma, J. P., Tucker, A. D., Leonard, K., Pattus, F. *et al.* Structure of the Aeromonas toxin proaerolysin in its water-soluble and membrane-channel states. *Nature* **367**, 292-295 (1994).

89. Stefureac, R., Long, Y. T., Kraatz, H. B., Howard, P. & Lee, J. S. Transport of alpha-helical peptides through alpha-hemolysin and aerolysin pores. *Biochemistry* **45**, 9172-9179 (2006).

90. Goodrich, C. P., Kirmizialtin, S., Huyghues-Despointes, B. M., Zhu, A., Scholtz, J. M., Makarov, D. E. *et al.* Single-molecule electrophoresis of beta-hairpin peptides by electrical recordings and Langevin dynamics simulations. *J Phys Chem B* **111**, 3332-3335 (2007).

91. Kang, X. F., Gu, L. Q., Cheley, S. & Bayley, H. Single protein pores containing molecular adapters at high temperatures. *Angew Chem Int Ed Engl* **44**, 1495-1499 (2005).

92. Japrung, D., Henricus, M., Li, Q., Maglia, G. & Bayley, H. Urea facilitates the translocation of single-stranded DNA and RNA through the alpha-hemolysin nanopore. *Biophys J* **98**, 1856-1863 (2010).

93. Pastoriza-Gallego, M., Rabah, L., Gibrat, G., Thiebot, B., van der Goot, F. G., Auvray, L. *et al.* Dynamics of unfolded protein transport through an aerolysin pore. *J Am Chem Soc* **133**, 2923-2931 (2011).

94. Merstorf, C., Cressiot, B., Pastoriza-Gallego, M., Oukhaled, A., Betton, J. M., Auvray, L. *et al.* Wild type, mutant protein unfolding and phase transition detected by single-nanopore recording. *ACS Chem Biol* **7**, 652-658 (2012).

95. Payet, L., Martinho, M., Pastoriza-Gallego, M., Betton, J. M., Auvray, L., Pelta, J. *et al.* Thermal unfolding of proteins probed at the single molecule level using nanopores. *Anal Chem* **84**, 4071-4076 (2012).

96. Nivala, J., Marks, D. B. & Akeson, M. Unfoldase-mediated protein translocation through an alpha-hemolysin nanopore. *Nat Biotechnol* **31**, 247-250 (2013).

97. Rodriguez-Larrea, D. & Bayley, H. Multistep protein unfolding during nanopore translocation. *Nat Nanotechnol* **8**, 288-295 (2013).

98. Aubin-Tam, M. E., Olivares, A. O., Sauer, R. T., Baker, T. A. & Lang, M. J. Single-molecule protein unfolding and translocation by an ATP-fueled proteolytic machine. *Cell* **145**, 257-267 (2011).

99. Maillard, R. A., Chistol, G., Sen, M., Righini, M., Tan, J., Kaiser, C. M. *et al.* ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell* **145**, 459-469 (2011).

100. Nivala, J., Mulroney, L., Li, G., Schreiber, J. & Akeson, M. Discrimination among protein variants using an unfoldase-coupled nanopore. *ACS Nano* **8**, 12365-12375 (2014).

101. Baker, T. A. & Sauer, R. T. ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochim Biophys Acta* **1823**, 15-28 (2012).

102. Thoren, K. L., Worden, E. J., Yassif, J. M. & Krantz, B. A. Lethal factor unfolding is the most force-dependent step of anthrax toxin translocation. *Proc Natl Acad Sci U S A* **106**, 21555-21560 (2009).

103. Basilio, D., Jennings-Antipov, L. D., Jakes, K. S. & Finkelstein, A. Trapping a translocating protein within the anthrax toxin channel: implications for the secondary structure of permeating proteins. *J Gen Physiol* **137**, 343-356 (2011).

104. Rabin, Y. & Tanaka, M. DNA in nanopores: counterion condensation and coion depletion. *Phys Rev Lett* **94**, 148103 (2005).

105. Zhang, J. & Shklovskii, B. I. Effective charge and free energy of DNA inside an ion channel. *Phys Rev E Stat Nonlin Soft Matter Phys* **75**, 021906 (2007).

106. Rodriguez-Larrea, D. & Bayley, H. Protein co-translocational unfolding depends on the direction of pulling. *Nat Commun* **5**, 4841 (2014).

107. Munoz, V. & Sanchez-Ruiz, J. M. Exploring protein-folding ensembles: a variable-barrier model for the analysis of equilibrium unfolding experiments. *Proc Natl Acad Sci U S A* **101**, 17646-17651 (2004).

108. Pey, A. L., Rodriguez-Larrea, D., Bomke, S., Dammers, S., Godoy-Ruiz, R., Garcia-Mira, M. M. *et al.* Engineering proteins with tunable thermodynamic and kinetic stabilities. *Proteins* **71**, 165-174 (2008).

109. Sato, T., Esaki, M., Fernandez, J. M. & Endo, T. Comparison of the protein-unfolding pathways between mitochondrial protein import and atomic-force microscopy measurements. *Proc Natl Acad Sci U S A* **102**, 17999-18004 (2005).

110. Berko, D., Tabachnick-Cherny, S., Shental-Bechor, D., Cascio, P., Mioletti, S., Levy, Y. *et al.* The direction of protein entry into the proteasome determines the variety of products and depends on the force needed to unfold its two termini. *Mol Cell* **48**, 601-611 (2012).

111. Rosen, C. B., Rodriguez-Larrea, D. & Bayley, H. Single-molecule site-specific detection of protein phosphorylation with a nanopore. *Nat Biotechnol* **32**, 179-181 (2014).

112. Ren, H., Xiao, F., Zhan, K., Kim, Y. P., Xie, H., Xia, Z. *et al.* A biocompatible condensation reaction for the labeling of terminal cysteine residues on proteins. *Angew Chem Int Ed Engl* **48**, 9658-9662 (2009).

113. Nguyen, D. P., Elliott, T., Holt, M., Muir, T. W. & Chin, J. W. Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation. *J Am Chem Soc* **133**, 11418-11421 (2011).

114. White, E. H., McCapra, F., Field, G. F. & McElroy, W. D. The structure and synthesis of firefly luciferin. *J Am Chem Soc* **83**, 2402-2403 (1961).

115. Liang, G., Ren, H. & Rao, J. A biocompatible condensation reaction for controlled assembly of nanostructures in living cells. *Nat Chem* **2**, 54-60 (2010).

116. Rodriguez-Larrea, D., Perez-Jimenez, R., Sanchez-Romero, I., Delgado-Delgado, A., Fernandez, J. M. & Sanchez-Ruiz, J. M. Role of conservative mutations in protein multi-property adaptation. *Biochem J* **429**, 243-249 (2010).

117. Gentle, I. E., De Souza, D. P. & Baca, M. Direct production of proteins with N-terminal cysteine for site-specific conjugation. *Bioconjug Chem* **15**, 658-663 (2004).

118. Matouschek, A. Protein unfolding--an important process in vivo? *Curr Opin Struct Biol* **13**, 98-109 (2003).

119. Prakash, S. & Matouschek, A. Protein unfolding in the cell. *Trends Biochem Sci* **29**, 593-600 (2004).

120. Lee, C. M., Sedman, J., Neupert, W. & Stuart, R. A. The DNA helicase, Hmi1p, is transported into mitochondria by a C-terminal cleavable targeting signal. *J Biol Chem* **274**, 20937-20942 (1999).

121. Huang, S., Ratliff, K. S. & Matouschek, A. Protein unfolding by the mitochondrial membrane potential. *Nat Struct Biol* **9**, 301-307 (2002).

122. van der Laan, M., Meinecke, M., Dudek, J., Hutu, D. P., Lind, M., Perschil, I. *et al.* Motor-free mitochondrial presequence translocase drives membrane integration of preproteins. *Nat Cell Biol* **9**, 1152-1159 (2007).

123. Krayl, M., Lim, J. H., Martin, F., Guiard, B. & Voos, W. A cooperative action of the ATP-dependent import motor complex and the inner membrane potential drives mitochondrial preprotein import. *Mol Cell Biol* **27**, 411-425 (2007).

124. Chacinska, A., Koehler, C. M., Milenkovic, D., Lithgow, T. & Pfanner, N. Importing mitochondrial proteins: machineries and mechanisms. *Cell* **138**, 628-644 (2009).

125. Paila, Y. D., Richardson, L. G. L. & Schnell, D. J. New insights into the mechanism of chloroplast protein import and its integration with protein quality control, organelle biogenesis and development. *J Mol Biol* **427**, 1038-1060 (2015).

126. Takeuchi, J., Chen, H. & Coffino, P. Proteasome substrate degradation requires association plus extended peptide. *EMBO J* **26**, 123-131 (2007).

127. Hagai, T. & Levy, Y. Ubiquitin not only serves as a tag but also assists degradation by inducing protein unfolding. *Proc Natl Acad Sci U S A* **107**, 2001-2006 (2010).

128. Martin, A., Baker, T. A. & Sauer, R. T. Pore loops of the AAA+ ClpX machine grip substrates to drive translocation and unfolding. *Nat Struct Mol Biol* **15**, 1147-1151 (2008).

129. Kenniston, J. A., Baker, T. A., Fernandez, J. M. & Sauer, R. T. Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell* **114**, 511-520 (2003).

130. Schlierf, M., Li, H. & Fernandez, J. M. The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proc Natl Acad Sci U S A* **101**, 7299-7304 (2004).

131. Dill, K. A. & MacCallum, J. L. The protein-folding problem, 50 years on. *Science* **338**, 1042-1046 (2012).

132. Gershenson, A. & Gierasch, L. M. Protein folding in the cell: challenges and progress. *Curr Opin Struct Biol* **21**, 32-41 (2011).

133. Gruebele, M., Dave, K. & Sukenik, S. Globular Protein Folding In Vitro and In Vivo. *Annu Rev Biophys* **45**, 233-251 (2016).

134. Balchin, D., Hayer-Hartl, M. & Hartl, F. U. In vivo aspects of protein folding and quality control. *Science* **353**, aac4354 (2016).

135. Nilsson, O. B., Hedman, R., Marino, J., Wickles, S., Bischoff, L., Johansson, M. *et al.* Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Rep* **12**, 1533-1540 (2015).

136. Holtkamp, W., Kokic, G., Jager, M., Mittelstaet, J., Komar, A. A. & Rodnina, M. V. Cotranslational protein folding on the ribosome monitored in real time. *Science* **350**, 1104-1107 (2015).

137. Wruck, F., Katranidis, A., Nierhaus, K. H., Buldt, G. & Hegner, M. Translation and folding of single proteins in real time. *Proc Natl Acad Sci U S A* **114**, E4399-E4407 (2017).

138. Kramer, G., Boehringer, D., Ban, N. & Bukau, B. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat Struct Mol Biol* **16**, 589-597 (2009).

139. Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M. & Hartl, F. U. Molecular chaperone functions in protein folding and proteostasis. *Annu Rev Biochem* **82**, 323-355 (2013).

140. Reid, D. W. & Nicchitta, C. V. Diversity and selectivity in mRNA translation on the endoplasmic reticulum. *Nat Rev Mol Cell Biol* **16**, 221-231 (2015).

141. Rapoport, T. A., Li, L. & Park, E. Structural and Mechanistic Insights into Protein Translocation. *Annu Rev Cell Dev Biol* **33**, 369-390 (2017).

142. Uversky, V. N., Oldfield, C. J. & Dunker, A. K. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys* **37**, 215-246 (2008).

143. Krishna, M. M., Hoang, L., Lin, Y. & Englander, S. W. Hydrogen exchange methods to study protein folding. *Methods* **34**, 51-64 (2004).

144. Englander, S. W. & Mayne, L. The nature of protein folding pathways. *Proc Natl Acad Sci U S A* **111**, 15873-15880 (2014).

145. Bartlett, A. I. & Radford, S. E. An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nat Struct Mol Biol* **16**, 582-588 (2009).

146. Michalet, X., Weiss, S. & Jager, M. Single-molecule fluorescence studies of protein folding and conformational dynamics. *Chem Rev* **106**, 1785-1813 (2006).

147. Schuler, B. & Eaton, W. A. Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol* **18**, 16-26 (2008).

148. Kapanidis, A. N. & Strick, T. Biology, one molecule at a time. *Trends Biochem Sci* **34**, 234-243 (2009).

149. Fisher, T. E., Marszalek, P. E. & Fernandez, J. M. Stretching single molecules into novel conformations using the atomic force microscope. *Nat Struct Biol* **7**, 719-724 (2000).

150. Moffitt, J. R., Chemla, Y. R., Smith, S. B. & Bustamante, C. Recent advances in optical tweezers. *Annu Rev Biochem* **77**, 205-228 (2008).

151. Borgia, A., Williams, P. M. & Clarke, J. Single-molecule studies of protein folding. *Annu Rev Biochem* **77**, 101-125 (2008).

152. Chivers, C. E., Crozat, E., Chu, C., Moy, V. T., Sherratt, D. J. & Howarth, M. A streptavidin variant with slower biotin dissociation and increased mechanostability. *Nat Methods* **7**, 391-393 (2010).

153. Sigworth, F. J. & Sine, S. M. Data transformations for improved display and fitting of single-channel dwell time histograms. *Biophys J* **52**, 1047-1054 (1987).

154. Colquhoun, D. Practical analysis of single channel records. in *Microelectrode Techniques: The Plymouth Workshop Handbook* (ed. Ogden, D. C.) 101-139 (Company of Biologists, Cambridge, UK, 1994).

155. Colquhoun, D. & Hawkes, A. G. The interpretation of single channel recordings. in *Microelectrode Techniques: The Plymouth Workshop Handbook* (ed. Ogden, D. C.) 141-188 (Company of Biologists, Cambridge, UK, 1994).

156. Colquhoun, D. & Hawkes, A. G. The principles of the stochastic interpretation of ion-channel mechanisms. in *Single-Channel Recording* (eds. Sakmann, B. & Neher, E.) 397-482 (Springer, Boston, MA, 1995).

157. Kelley, R. F. & Richards, F. M. Replacement of proline-76 with alanine eliminates the slowest kinetic phase in thioredoxin folding. *Biochemistry* **26**, 6765-6774 (1987).

158. Roderer, D. J., Scharer, M. A., Rubini, M. & Glockshuber, R. Acceleration of protein folding by four orders of magnitude through a single amino acid substitution. *Sci Rep* **5**, 11840 (2015).

159. Di Marino, D., Bonome, E. L., Tramontano, A. & Chinappi, M. All-Atom Molecular Dynamics Simulation of Protein Translocation through an alpha-Hemolysin Nanopore. *J Phys Chem Lett* **6**, 2963-2968 (2015).

160. Rubini, M., Scharer, M. A., Capitani, G. & Glockshuber, R. (4R)- and (4S)-fluoroproline in the conserved cis-prolyl peptide bond of the thioredoxin fold: tertiary structure context dictates ring puckering. *Chembiochem* **14**, 1053-1057 (2013).

161. Stefureac, R., Waldner, L., Howard, P. & Lee, J. S. Nanopore analysis of a small 86-residue protein. *Small* **4**, 59-63 (2008).

162. Guinn, E. J., Jagannathan, B. & Marqusee, S. Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein. *Nat Commun* **6**, 6861 (2015).

163. Schonfelder, J., Perez-Jimenez, R. & Munoz, V. A simple two-state protein unfolds mechanically via multiple heterogeneous pathways at single-molecule resolution. *Nat Commun* **7**, 11777 (2016).

164. Bayley, H., Luchian, T., Shin, S.-H. & Steffensen, M. B. Single-molecule covalent chemistry in a protein nanoreactor. in *Single Molecules and Nanotechnology* (eds. Rigler, R. & Vogel, H.) 251-277 (Springer, Berlin, Heidelberg, 2008).

165. Tian, P. & Andricioaei, I. Repetitive pulling catalyzes co-translocational unfolding of barnase during import through a mitochondrial pore. *Journal of Molecular Biology* **350**, 1017-1034 (2005).

166. Howorka, S., Cheley, S. & Bayley, H. Sequence-specific detection of individual DNA strands using engineered nanopores. *Nat Biotechnol* **19**, 636-639 (2001).

167. Cheley, S., Braha, O., Lu, X., Conlan, S. & Bayley, H. A functional protein pore with a "retro" transmembrane domain. *Protein Sci* **8**, 1257-1267 (1999).

168. Braha, O., Walker, B., Cheley, S., Kasianowicz, J. J., Song, L. Z., Gouaux, J. E. *et al.* Designed protein pores as components for biosensors. *Chemistry & Biology* **4**, 497-505 (1997).

169. Mayer, M., Kriebel, J. K., Tosteson, M. T. & Whitesides, G. M. Microfabricated teflon membranes for low-noise recordings of ion channels in planar lipid bilayers. *Biophys J* **85**, 2684-2695 (2003).

170. Montal, M. & Mueller, P. Formation of bimolecular membranes from lipid monolayers and a study of their electrical properties. *Proc Natl Acad Sci U S A* **69**, 3561-3566 (1972).

171. Movileanu, L., Cheley, S., Howorka, S., Braha, O. & Bayley, H. Location of a constriction in the lumen of a transmembrane pore by targeted covalent attachment of polymer molecules. *J Gen Physiol* **117**, 239-252 (2001).

172. Maglia, G., Heron, A. J., Stoddart, D., Japrung, D. & Bayley, H. Analysis of single nucleic acid molecules with protein nanopores. *Methods Enzymol* **475**, 591-623 (2010).